Observational cosmology with galaxy surveys

Autor: David Alonso Monge

Director: Juan García-Bellido Capdevila

Memoria de tesis doctoral presentada ante el Departamento de Física Teórica de la Universidad Autónoma de Madrid para la obtención del título de Doctor en Ciencias Físicas - Programa de Doctorado en Física Teórica.

May 14, 2013

Instituto de Física Teórica UAM-CSIC

A mi familia

Contents

Statement of authorship ix

Resumen y conclusiones xi

Acknowledgements xv

Notation and conventions xvii

Preface xxiii

1 The standard model of cosmology 1
- 1.1 The FRW metric . . . 2
  - 1.1.1 Cosmological parameters . . . 2
  - 1.1.2 Redshift and distances . . . 4
  - 1.1.3 Brief history of the Universe . . . 6
- 1.2 Late-time acceleration and the problem of dark energy . . . 7
  - 1.2.1 Cosmological constant . . . 8
  - 1.2.2 Alternative dark energy models . . . 9
  - 1.2.3 Modified gravity models . . . 10
  - 1.2.4 Accelerated expansion without DE . . . 10

2 Cosmological structure formation 13
- 2.1 Newtonian linear perturbation theory . . . 14
  - 2.1.1 Density perturbations . . . 15
  - 2.1.2 The velocity field . . . 16
  - 2.1.3 Redshift-space distortions . . . 17
- 2.2 From the drag epoch to the late-time power spectrum . . . 20
  - 2.2.1 Physics of perturbations before the drag epoch . . . 21
  - 2.2.2 Transfer functions and the CDM model . . . 23
  - 2.2.3 Fitting functions and baryonic features . . . 24
- 2.3 Non-linearities I: analytic and phenomenological approaches . . . 25
  - 2.3.1 Higher-order corrections . . . 25
  - 2.3.2 Lagrangian perturbation theory . . . 27
  - 2.3.3 The halo model . . . 29
- 2.4 Non-linearities II: numerical simulations . . . 37
  - 2.4.1 Initial conditions . . . 38
  - 2.4.2 Codes . . . 38

3 Model-independent determination of the BAO scale in redshift surveys 41
3.1 Standard approach to baryon acoustic oscillations . . . 41
3.1.1 Galaxy redshift surveys . . . 42
3.1.2 Measuring the BAO from the correlation function monopole . . . 43
3.1.3 Statistical uncertainties and the covariance matrix . . . 45
3.1.4 The effect of non-linearities . . . 46
3.2 Angular BAO in photometric redshift surveys . . . 47
3.2.1 Modelling angular clustering . . . 47
3.2.2 The angular sound horizon scale . . . 48
3.2.3 Method to recover $\theta_{\rm BAO}$ . . . 49
3.2.4 Applicability to the Dark Energy Survey . . . 54
3.2.5 Discussion . . . 61
3.3 Measuring the radial BAO scale . . . 62
3.3.1 Galaxy clustering in the radial direction . . . 63
3.3.2 Method to measure the radial BAO scale . . . 65
3.3.3 Application to a simulated galaxy survey . . . 67
3.3.4 Discussion . . . 73
3.4 Combining angular and radial BAO information . . . 74
3.4.1 Combined angular and radial BAO measurements . . . 74
3.4.2 Comparison with the standard analysis . . . 76
3.4.3 Discussion . . . 78

4 Structure formation in LTB models 81
4.1 Lemaître-Tolman-Bondi models . . . 82
4.1.1 The LTB metric . . . 82
4.1.2 Redshift . . . 83
4.1.3 Distances . . . 83
4.1.4 Einstein's equations . . . 84
4.1.5 Void models . . . 86
4.1.6 Perturbation theory in LTB . . . 87
4.2 Observational tests of LTB models . . . 88
4.2.1 Contrast with present observations . . . 88
4.2.2 Proposed observables and future experiments . . . 91
4.2.3 Fine-tuning in void models . . . 93
4.3 Large-scale structure simulations of LTB models . . . 94
4.3.1 Simulating voids . . . 95
4.3.2 The simulations . . . 95
4.3.3 Background evolution . . . 97
4.3.4 Density contrast evolution . . . 102
4.3.5 Discussion . . . 102
4.4 Halo statistics in LTB models . . . 104
4.4.1 The mass function in LTB models . . . 105
4.4.2 The halo catalogue . . . 106
4.4.3 Halo abundances . . . 106
4.4.4 Halo orientations . . . 109
4.4.5 Discussion . . . 111

5 Conclusions 113

Appendix A Statistics of the matter distribution **117**
A.1 Random fields . . . 117
A.1.1 Definitions and important results . . . 117
A.1.2 Filtering . . . 119
A.1.3 The two-point correlation function and the power spectrum . . . 119
A.1.4 Angular correlations functions and C_l's . . . 120
A.1.5 Redshift distortions and correlation functions . . . 123
A.2 Discrete tracers of the matter distribution . . . 124
A.2.1 Point processes . . . 125
A.2.2 Estimating the correlation function from discrete data . . . 126

Appendix B Gaussian covariance matrices for galaxy correlation functions **129**
B.1 Correlation function in real space . . . 129
B.2 Redshift space in the (π, σ) frame . . . 131
B.2.1 Particular case: the radial correlation function . . . 131
B.3 Redshift space in the (r, μ) frame . . . 132
B.3.1 Special case: the monopole . . . 133
B.4 Angular covariance . . . 133

Appendix C Geodesics and curvature tensors of the LTB metric **137**
C.1 Geodesics . . . 137
C.2 Curvature . . . 138

Statement of authorship

This thesis contains results based on original work done by myself in collaboration with other researchers during my Ph.D. from October 2009 to May 2013. The original results can be classified into two different topics, both related to the study of the accelerated expansion of the Universe and the problem of dark energy:

- **Measurement of the Baryon Acoustic Oscillations in redshift surveys.** Chapter 3 contains the results regarding references 1 and 2 listed below. These are detailed in sections 3.2, 3.3 and 3.4, while section 3.1 recopilates our current knowledge in the field (therefore no claim is made about this section).
- **Structure formation in LTB models.** Chapter 4 contains the results published in the references 3 and 4 listed below. Sections 4.1 and 4.2 cover the current knowledge about this topic, and the original results derived from the aforementioned references are detailed in sections 4.3 and 4.4.

Chapters 1 and 2 are review chapters, based on the literature I have read during my Ph.D. studentship, and that has shaped my knowledge of cosmology. No new result is presented in these chapters, and no originality is claimed. Appendices A, B and C present lengthy calculations and results that are relevant for the work presented here.

David Alonso Monge
May, 2013

Research papers

1. Sánchez E., Carnero A., García-Bellido J., Gaztañaga E., de Simoni F., Crocce M., Cabre A., Fosalba P., Alonso D. (2011). MNRAS, **411**:277. arXiv:1006.3226
2. Sánchez E., Alonso D., Sánchez F. J., García-Bellido J., Sevilla I. (2012). arXiv:1210.6446 (submitted to MNRAS)
3. Alonso D., García-Bellido J., Haugbøelle T., Vicente J. (2010). PRD, **82**:123530. arXiv:1010.3453
4. Alonso D., García-Bellido J., Haugbølle T., Knebe A. (2012). Phys.Dark Univ. **1**:24. arXiv:1204.3532v2

Resumen y conclusiones

La cosmología trata de responder a algunas de las cuestiones más fundamentales que el ser humano siempre se ha preguntado: cuál es el origen y el destino del Universo. Sin embargo, no hace mucho que podemos decir que la cosmología se encuentra entre las ciencias empiricas, alejada de la especulación y el misticismo, capaz de realizar predicciones concretas y basada en observaciones objetivas. Esto no es sorprendente: toda ciencia que se precie debe basarse en la observación de la naturaleza y la experimentación, y responder a las preguntas más fundamentales requiere las observaciones más sofisticadas. Hoy en día, sin embargo, disponemos de multitud de datos observacionales que permiten hacernos una idea de cómo es el Universo en el que nos ha tocado vivir: las abundancias relativas de elementos ligeros, las observaciones del fondo de radiación de microondas (CMB), la atenuación de la luz proviniente de supernovas o los estudios sobre formación de estructura a gran escala. Los resultados presentados recientemente por la colaboración Planck reflejan claramente los avances que han tenido lugar en este campo. Ahora somos capaces de determinar los parámetros de un modelo cosmológico estándar, que parecen regir la evolución del Universo, con precisiones del orden del $\sim 2\%$.

Sin embargo, el hecho de que conozcamos el valor de estos parámetros no implica que entendamos el Universo que parecen describir. Según el modelo cosmológico actual, tan sólo el 20% de toda la materia del Universo está compuesta de materia bariónica ordinaria, que podemos entender de acuerdo con el modelo estándar de física de partículas. El otro 80% se encuentra compuesto por lo que se denomina *materia oscura*, un tipo de materia que no interacciona electromagnéticamente, y que podemos observar sólo a través de sus efectos gravitacionales. Esta situación es bastante desconcertante, y se convierte casi en insostenible si analizamos las distintas contribuciones a la densidad de energía total del Universo. La observación de la expansión acelerada del Universo parece indicar que debe existir una forma de energía que actua como una fuente gravitatoria repulsiva. Esta energía oscura constituye en torno al 70% de la densidad de energía total, lo que implica que solamente tenemos un modelo basado en primeros principios capaz de describir el 5% del contenido energético del Universo.

En la tésis expuesta en este documento nos hemos centrado en ese 70%: en el problema de la energía oscura y la expansión acelerada del Universo. Específicamente, esta tésis cubre dos aspectos concretos, que describimos a continuación.

La medida de la escala de bariones en catálogos de galaxias

En el Universo primitivo, fotones y bariones se encuentran fuertemente acoplados formando un único fluido imperfecto. Los gradientes de presión en este fluido previenen el colapso gravitacional y generan ondas sonoras hasta el momento del desacoplo de los fotones. Este

proceso deja su impronta en la distribución de materia a gran escala que podemos observar hoy en día. En concreto, existe una escala característica, el horizonte de sonido r_s, que corresponde a la distancia máxima que una onda sonora ha podido viajar hasta el momento en que los bariones dejan de sentir la interacción con la radiación. Esta escala puede ser observada en la función de correlación de materia como un exceso de probabilidad de encontrar dos objetos a una distancia r_s. Al ser el horizonte sonoro una escala estándar, puede ser utilizada para determinar la relación redshift-distancia, lo que da información clave acerca de la expansión del Universo. Más precisamente, estas oscilaciones acústicas de bariones, normalmente conocidas por sus siglas en inglés BAO (Baryon Acoustic Oscillations), pueden proporcionarnos información a través de dos observables independientes: la escala de BAO angular, $\theta_{\rm BAO}$, y la escala de bariones a lo largo de la linea de visión, o BAO radial, $\Delta z_{\rm BAO}$. Mientras que la escala angular puede usarse para determinar la dependencia de la distancia diámetro angular con el redshift, la escala radial proporciona una medida directa del ritmo de expansión.

A pesar de la existencia de estos dos observables independientes, la mayoría de las detecciones de las BAO realizadas hasta la fecha han medido tan sólo una combinación de las dos, formada por la media geométrica de ambas. Esta medida se realiza sobre el monopolo de la función de correlación tridimensional, calculada como una media sobre las distintas orientaciones de los pares de galaxias con respecto a la linea de visión. Esta es la forma más eficiente de utilizar los datos en terminos de estadística, dado que se utilizan todos los pares de galaxias. Sin embargo, los datos de futuros catálogos de galaxias como DES-spec, BigBOSS o Euclid serán de una calidad suficientemente buena como para que no sea necesario un uso tan eficiente, sino que las dos escalas de BAO puedan ser determinadas de manera independiente. Otra desventaja del método estándar es el uso de la función de correlación 3D, dado que en su estimación es necesario calcular las distancias tridimensionales entre galaxias. Para ello hace falta determinar la distancia a cada galaxia a partir de su corrimiento al rojo (redshift), usando un modelo cosmológico de referencia. No está claro que al hacer esto no introduzcamos un sesgo en los resultados finales del análisis hacia esta cosmología de referencia. Si bien es cierto que los efectos debidos a la diferencia entre la cosmología real y la de referencia pueden modelizarse cuando éstas son parecidas de acuerdo con el llamado efecto Alcock-Paczynski, no existe una medida clara del concepto de "cercanía" entre dos modelos cosmológicos y no hay ningún estudio concluyente acerca de los efectos de esta elección en los resultados finales de las medidas de las BAO. Podemos hacernos una idea acerca del modelo cosmológico verdadero a partir de observables provinientes del Universo primitivo (p.ej. el CMB) y de observables a bajo redshift, como supernovas, sin embargo debemos ser muy cuidadosos antes de usar un modelo cosmológico concreto para analizar los datos que en el futuro obtendremos en el rango inexplorado de redshifts intermedios a partir de catálogos de galaxias o de medidas de las lineas de emisión Lyman-α y 21cm.

Por estas razones, en las referencias [1, 2] hemos propuesto un método alternativo para medir la escala del horizonte sonoro en las direcciones angular y radial de forma independiente. Este método utiliza observables puros (redshifts y ángulos) y sus resultados son, por lo tanto, independientes del modelo cosmológico. El análisis está basado en el ajuste de las funciones de correlación angular y radial a un modelo empírico que consiste en una función suave que describe la forma de la función de correlación antes y después de la escala de BAO, y un término Gaussiano que determina la posición de la escala de bariones.

Para medir la función de correlación angular es necesario proyectar la distribución de galaxias en intervalos de redshift relativamente anchos. Esto afecta al pico de BAO, haciéndolo más ancho y desplazándolo ligeramente. Sin embargo, como demostramos en esta tesis, estos efectos de proyección pueden ser incorporados al método sin disminuir su sensitividad. Este

método puede también ser utilizado para determinar la escala angular de bariones a partir de catálogos fotométricos de galáxias. Sin embargo, el uso de redshifts fotométricos, de menor precisión que los espectroscópicos, introduce una fuente dominante de errores sistemáticos.

En el caso de la función de correlación radial, los efectos de proyección se limitan a escalas pequeñas y no afectan a la posición del pico. Sin embargo, la detección del BAO radial se encuentra limitada por el volumen explorado por el catálogo, dado que la varianza cósmica juega un papel dominante en esta medida. Esto no es sorprendente, dado que sólo una fracción pequeña de los pares de galaxias pueden estar orientados a lo largo de la dirección radial.

En definitiva, el método es extremadamente robusto frente a errores sistemáticos (escepto en el caso fotométrico) y puede usarse para obtener una medida precisa de la escala de BAO. Hemos probado la efectividad del método en un catálogo simulado de galaxias proporcionado por la colaboración MICE, y hemos verificado que se pueden obtener cotas bastante restrictivas sobre los parámetros cosmológicos usando tan sólo datos de BAO cuando se combinan medidas angulares y radiales. Además hemos comparado estos resultados con los obtenidos a partir del análisis estándar del mismo catálogo. Ambos métodos parecen dar resultados comparables, con la ventaja añadida de que el método propuesto no necesita tomar ninguna cosmología de referencia, un hecho que puede ser crucial a la hora de poner cotas sobre modelos alternativos de energía oscura.

Formación de estructura en modelos LTB

Para asegurarnos de que los datos observacionales indican la existencia de fenómenos no predichos por el modelo teórico actual, es necesario en primer lugar asegurarse de que todas las suposiciones hechas a la hora de analizar e interpretar dichos datos son ciertas. Uno de los pilares de la cosmología actual es el llamado Principio Cosmológico, según el cual, el Universo que observamos a nuestro alrededor debe ser estadísticamente equivalente al observado desde cualquier otro punto del espacio. La validez de este principio es difícil de verificar, dado que nos es imposible viajar distancias cosmológicas para medir las propiedades estadísticas del Universo desde otro punto. Por esta razón distintos grupos han explorado las consecuencias de prescindir del Principio Cosmológico. Una posibilidad es que nos encontrásemos cerca del centro de una gran región esféricamente simétrica de baja densidad (un vacío). Se ha señalado que en esta situación, las observaciones astronómicas interpretadas en términos de una cosmología homogénea podrían confundirse con una expansión acelerada del Universo. Estos modelos se denominan modelos de vacío o modelos LTB, dado que su descripción matemática se realiza en términos de la métrica de Lemaître-Tolman-Bondi. A pesar de que los modelos LTB implican una violación explícita del Principio Copernicano, es un sano ejercicio científico tratar de descartarlos utilizando tan sólo datos observacionales. Durante los últimos años este objetivo ha generado una gran actividad científica y hoy podemos afirmar que la gran mayoría de los modelos LTB son incompatibles con las observaciones.

Nuestra contribución al estudio de los modelos LTB ha sido el desarrollo de una técnica para realizar simulaciones numéricas de formación de estructura en dichos modelos. Ésto nos permitiría comprender mejor la evolución de las perturbaciones en la densidad de materia en un entorno esféricamente simétrico, algo áltamente no trivial incluso en el régimen lineal, dado que la teoría de perturbaciones sobre la métrica LTB es bastante más complicada que en el caso estándar. La técnica utilizada se basa en comprender que el vacío LTB se genera a partir de una perturbación inicial de gran escala pero baja amplitud. El colapso gravitacional hace que esta perturbación crezca, evolucionando hacia valores no lineales. Dado que los modelos LTB no incluyen ningún elemento de nueva física (el único componente de materia

es materia no relativista y la gravedad sigue estando regida por la Relatividad General), sólo necesitamos generar las condiciones iniciales de la simulación de manera apropiada, para que se genere un vacío con las propiedades deseadas. Dado que conocemos las soluciones exactas para la métrica LTB, esta tarea es relativamente sencilla, y puede llevarse a cabo modificando un generador de condiciones iniciales para que tenga en cuenta tanto las perturbaciones a pequeña escala debidas a las fluctuaciones primordiales como la debida al vacío a gran escala.

Los parámetros de estas simulaciones deben elegirse con cuidado. Debido a que las simulaciones de formación de estructura utilizan una caja con condiciones de borde periódicas, al utilizar esta técnica no estaremos simulando símplemente un vacío LTB, sino una red de vacíos formada por sus repeticiones periódicas. Para poder despreciar la interacción entre los vacíos de la red es necesario utilizar una caja de tamaño considerablemente mayor que el diámetro del vacío. Hemos verificado que al hacer esto los efectos de borde desaparecen completamente. Dado que los vacíos viables como modelos alternativos a la energía oscura son enormes (en torno a 2-3 Gpc), la caja de la simulación debe tener un tamaño considerable. Esto hace que sea necesario simular un gran número de partículas para poder resolver estructuras pequeñas, con lo que llevar a cabo una de estas simulaciones puede ser computacionalmente muy costoso. Hemos verificado que nuestras simulaciones consiguen seguir las soluciones exactas de LTB tanto en densidades como en velocidades, para un amplio rango de modelos.

Una de las características de los modelos LTB es el hecho de que, a un radio comóvil fijo, evolucionan de manera equivalente a un modelo homogéneo FRW (Friedmann-Robertson-Walker) con unos parámetros cosmológicos efectivos. Si bien esto es cierto a orden 0, las perturbaciones no evolucionan de manera análoga en ambos casos. El hecho de que estos modelos tengan dos ritmos de expansión distintos a lo largo de las direcciones radial y transversal implica que los objetos extensos no sólo se expanden en volumen, sino que también cambian de forma. Esto afecta al modo en que las perturbaciones crecen y hace que este crecimiento sea diferente al del modelo FRW efectivo. Hemos utilizado nuestras simulaciones numéricas para estudiar este efecto. Inicialmente extrajimos el crecimiento de estructura a partir de la varianza del contraste de densidad a distintas distancias del centro del vacío. Verificamos que el crecimiento era compatible con el de un Universo abierto en el interior del vacío, mientras que en el exterior se comportaba como lo haría en un Universo plano Einstein-deSitter. Este resultado fue explorado de manera cuantitativa estudiando la abundancia de halos de materia oscura de diferente masa en nuestras simulaciones. El resultado principal de este estudio es que las abundancias de halos en los modelos LTB no son compatibles con la evolución de las perturbaciones en el universo FRW efectivo y que es necesario incluir una corrección proporcional a la diferencia normalizada entre los ritmos de expansión radial y transversal (que denominamos *background shear*). Hemos comprobado que esta corrección es constante a distintos radios en la simulación utilizada para este análisis. Si además fuese independiente de los parámetros del modelo LTB, las abundancias de objetos virializados podrían utilizarse para detectar la presencia de un *background shear*, un claro indicador de una expansión inhomogénea.

Acknowledgements

Compiling a list of people who deserve one's gratitude after completing a Ph.D. thesis is similar to compiling the bibliography for that thesis: quite probably you will forget some reference, and someone will get angry because you did not include them. For this reason I would like to apologize in advance to all those who eagerly and unsuccessfully look for their names in the list below. I can a assure you that there is a high chance that the unforgivable omission is due to my own absent-mindedness. If after this apology you still don't think I deserve any forgiveness for my dispeakable transgression, I would suggest that stealing my keys and hiding them around could be a good punishment. That drives me mad.

The following list of acknowledgements is separated into three categories. The order of these and of the people and institutions mentioned within them is mainly arbitrary and not related to their impact on the work presented here.

Scientific acknowledgements

First and foremost, my supervisor Juan García-Bellido deserves my most sincere gratitude for his support throughout these years, as well as for the many opportunities that, without him, I probably wouldn't have had access to. Working with Juan has allowed me to understand many of the qualities that are necessary in a good scientist.

The CIEMAT crew (Eusebio, Nacho, Javi y Rafa) also deserve a very important place in this list. Thanks to them I was able to learn a great deal about the real world, which an annoying theoritician can only glimpse behind Lagrangians and manifolds. I haven't forgotten the many beers I still owe you!

During my studentship I had the privilege of paying a 4-month visit in 2011 to the Institute for Astronomy, at the Royal Observatory of Edinburgh. This visit had an important impact on my understanding of cosmology and therefore also on the contents of this work. This opportunity was granted to me in a great deal thanks to John A. Peacock. I would like to take the chance to thank him for this and for the support that he has shown me since then. All the other kind people at the observatory (Paula, Neil and Edo in particular) should also be acknowledged for making my stay an extremely pleasant one.

My first scientific publication was carried out in collaboration with Troels Haugbølle. His enthusiasm about science and his many useful tips have been an invaluable contribution to my scientific education.

I am also indebted to Alexander Knebe, the "dog guy" at UAM, for his insight during our collaboration and for his later support.

My scientific siblings (Dani, Zuma, Alicia and Santi) have obviously played an important role during the last few years, and their voluntary and involuntary impact should be acknowl-

edged here. Together with them I must also thank Federico Urban and Lluis Galbany, two very important "cosmolocos". May we enjoy more and bigger chuletones in the years to come!

Formal acknowledgements

It would have never been possible for me to pursue a career as a scientific researcher and to contribute to the collective understanding of nature if it weren't for the Spanish public education system, which is now in grave danger due to the economical policies enforced by the current and previous governments. In particular I would like to acknowledge the work done at the Universidad Autónoma de Madrid, which has offered me many generous opportunities since 2005.

I acknowledge support from a JAE-Predoc scholarship, granted by The Spanish High Council of Scientific Research (CSIC), thanks to which I have been able to work on my Ph.D. thesis.

A significant part of the work presented in this thesis was carried out thanks to the simulated data kindly provided by the MICE collaboration [1].

Family and friends

A mis padres, Julia y Jorge, a mis hermanos Ester y Raúl, a mi abuelo Honorio y a mi tía Pilar debo agradecer, más que a nadie, ser la persona que soy hoy, ya que son los valores que he aprendido de cada uno de ellos los que me han permitido guiarme en la vida. A ellos dedico esta tesis y a ellos estaré siempre agradecido.

Para hacer un doctorado, primero hay que hacer una licenciatura, y sin los amigos que conocí en la Licenciatura de Física, aquellos cinco años habrían sido muy diferentes. A Laisfer, Pedro, Iván, Nacho y Julián (y a todos los demás que me dejo en el tintero) les agradezco haber estado tan mal de la cabeza como yo cuando decidieron que la Física podía ser un negocio lucrativo. A Lucy, que también compartió nuestra insensatez, le agracezco haber sembrado el sentido común necesario para que no acabásemos asesinándonos los unos a los otros.

Y para hacer una licenciatura, antes hay que ir al colegio y al instituto. Cuatro personas han influido decisivamente en mi vida antes de empezar la Universidad y ninguna lista de agradecimientos estaría completa sin incluirlos. Inma, Vichi, Jaime y Raúl: gracias por todo!

A Paloma le agradezco haber sufrido en silencio a un estudiante de doctorado escribiendo la tesis en casa las 24 horas al día, y todo el apoyo que me ha dado en los últimos años. Mi vida sin tí sería mucho menos entretenida.

¡GRACIAS A TODOS!

[1] http://maia.ice.cat/mice/

Notation and conventions

The following conventions have been adopted throughout this work unless otherwise stated in the text:

- In general, partial derivatives are written as $\partial_x \equiv \partial/\partial x$. When dealing with several coordinates labelled x_i, $i = 1, ..., N$ we shall also write $\partial_i \equiv \partial/\partial x_i$.
- The Dirac and Kronecker delta functions are $\delta^{\mathcal{D}}(x-y)$ and $\delta^{\mathcal{K}}_{x,y}$ respectively. When their arguments are vectors, their multi-dimensional generalizations are understood.
- The Fourier transform is defined as

$$f_{\mathbf{k}} = \int \frac{dk^3}{(2\pi)^{3/2}} f(\mathbf{x})\, e^{-i\,\mathbf{k}\cdot\mathbf{x}}, \tag{1}$$

  unless otherwise stated. One such exception is the relation between the power-spectrum and the correlation function

$$\xi(\mathbf{r}) = \frac{1}{(2\pi)^3} \int dk^3\, P(\mathbf{k})\, e^{i\,\mathbf{k}\cdot\mathbf{r}}. \tag{2}$$

- The metric signature is $(-,+,+,+)$.
- Quantities in real space and redshift space will be labelled with an r and an s subscript respectively whenever they could be mistaken.

List of acronyms and symbols

Term	**Description**
2PCF	2-point correlation function.
AU	astronomical unit: 1.496×10^8 km.
BAO	baryon acoustic oscillations.
CDM	cold dark matter.
CMB	Cosmic Microwave Background.
CP	Cosmological Principle.
DE	dark energy.
EdS	Einstein-deSitter.
FRW	Friedmann-Robertson-Walker.
GR	General Relativity.
H_0	Hubble expansion parameter.
h	$H_0/(100\,\mathrm{km\,s^{-1}\,Mpc^{-1}})$.
HOD	halo occupation distribution.
kSZ	kinematic Suynaev-Zel'dovich effect.
ΛCDM	Λ - Cold Dark Matter.
$L_l(x)$	Legendre polynomial of order l.
LPT	Lagrangian perturbation theory.
LSS	large-scale structure.
LTB	Lemaître-Tolman-Bondi.
$M_\odot$	Solar mass: 1.989×10^{30} kg.
w_a	Evolution of the DE equation of state.
n_s	Scalar spectral index.
Ω_x	Density parameter for the x component.
pc	parsec: $3.086 \times 10^1 3$ km.
photo-z	photometric redshift.
PS	Press & Schechter.
PT	perturbation theory.
QFT	Quantum Field Theory.
r.m.s.	root mean square.
RPT	renormalized perturbation theory.
RSD	redshift-space distortions.
σ_8	Matter power spectrum normalization.
SNe	supernovae.
SNeIa	type I-a supernovae.
w_0, w	Dark energy equation of state.

List of Figures

1.1 Summary of current BAO measurements 8

2.1 Illustrative representation of RSDs 19
2.2 Correlation function multipoles 20
2.3 Excursion-set formalism 32

3.1 The effect of non-linearities and projection effects on the angular correlation function 48
3.2 Parametrization of the angular correlation function 50
3.3 Shift $\theta_{\rm FIT} - \theta_{\rm BAO}$ due to projection effects 52
3.4 Systematic error associated to non-linearities 53
3.5 Effect of RSDs and photo-z on the angular correlation function 53
3.6 Convergence of real and redshift space for wide redshift bins 54
3.7 Photo-z precision used for the analysis in section 3.2 55
3.8 Correlation functions measured from the MICE mock catalogue and used in the analysis in section 3.2 56
3.9 Evolution of $\theta_{\rm BAO}$ measured from the MICE catalogue 57
3.10 Cosmological constraints drawn from the angular BAO of the MICE catalogue in section 3.2 58
3.11 Correlation matrix between redshift bins due to the photo-z 58
3.12 Dependence of $\theta_{\rm BAO}$ on the fit limits 60
3.13 Systematic error on $\theta_{\rm BAO}$ associated to the photo-z 61
3.14 Radial correlation function in real and redshift space 63
3.15 Fit to the theoretical radial correlation function 66
3.16 $N(z)$ in the MICE catalogue 67
3.17 Correlation functions measured from the MICE catalogue and used in the analysis in section 3.3 68
3.18 Different contributions to the total statistical error of the radial correlation function 69
3.19 Evolution of the radial BAO scale measured from the MICE catalogue 69
3.20 Angular projection effects on the radial correlation function 70
3.21 Variation of $\Delta z_{\rm BAO}$ with angular pixel size 70
3.22 Systematic error on $\Delta z_{\rm BAO}$ due to the fit limits 72
3.23 Cosmological constraints from the measurements of the radial BAO in the MICE catalogue 73

3.24 Cosmological constraints due to the combination of radial and angular BAO information . . . 75
3.25 Constraints including radial BAO, angular BAO and CMB data . . . 75
3.26 Monopole 2PCFs calculated from the MICE catalogue . . . 76
3.27 Measurements of the scaling parameter α from the MICE catalogue . . . 77
3.28 Comparison of the cosmological constraints obtained with the standard method to measure the BAO and the one proposed in this work . . . 78

4.1 LTB models and SNe . . . 89
4.2 LTB models and the kSZ effect . . . 89
4.3 Normalized background shear in void models . . . 92
4.4 Snapshots from N-body simulations of LTB models . . . 96
4.5 r-dependent curvature in void models . . . 98
4.6 NGP, CIC and TSC gridding algorithms . . . 99
4.7 Evolution of the density profile in N-body simulations of LTB models . . . 100
4.8 Density profiles for different LTB models . . . 100
4.9 Velocity profiles in LTB models . . . 101
4.10 Evolution of the density contrast in LTB models . . . 103
4.11 Density contrast at different r in LTB models . . . 104
4.12 Mass function in LTB models and background shear . . . 107
4.13 Goodness of fit of the shear correction for the halo mass function . . . 108
4.14 Halo orientations in LTB models . . . 110

A.1 The (r, μ) and (π, σ) coordinates for the anisotropic 2PCF . . . 123
A.2 Boundary effects in 2PCFs . . . 127

B.1 Covariance matrices . . . 135

List of Tables

3.1 Cosmological models used in section 3.2 50
3.2 Systematic errors associated to the measurement of the angular BAO scale . . 59

4.1 N-body simulations of LTB models 97
4.2 Goodness of fit of the shear correction to the mass function in LTB models . . 109
4.3 Goodness of fit of the shear correction at different redshifts 109

A.1 Estimators for the 2PCF from discrete data 128

Preface

Isn't it enough to see that a garden is beautiful without having to believe that there are fairies at the bottom of it too?
Douglas Adams [3]

Even though cosmology tries to give an answer to some of the most intriguing and fundamental questions mankind has been asking from time immemorial, only recently has this discipline been able to emerge from the bowels of superstition, mysticism and speculation to become a complete science, whose predictions can be tested against empirical data. It is no wonder that the most fundamental questions require the most sophisticated experimental methods.

Nowadays there exist numerous observational sources from which we can begin to understand the Universe we happen to live in: the relative abundances of light elements, the spectrum of the cosmic microwave background, galaxy cluster number counts, the luminosity distance of type-Ia supernovae, large-scale structure, etc., and many other observables show an extraordinary potential for the near future, such as measurements of the Lyman-α forest or the 21cm line of atomic hydrogen. The latest results presented by the Planck collaboration [4] are clear evidence that our knowledge of the Universe has improved since the speculative static universe of Einstein. We can now measure the values of the cosmological parameters with a precision of a few percent.

However, being able to measure the main components of the Universe does not mean that we understand the picture these quantities draw. First of all, baryonic matter, the kind that we are used to and that we can manipulate in a laboratory, seems to make up only 20% of the total amount of matter in the Universe. According to the standard cosmological model, the remaining 80% is in the form of dark matter: a component that does not feel the electromagnetic interaction and thus can only be observed through its gravitational effects. This component can not be described within the standard model of particle physics, and it has not been directly detected otherwise. This uncomfortable picture is made even worse if one considers the different contributions to the total energy density of the Universe. The observations of the accelerated expansion of the Universe seem to point towards the existence of a form of energy that acts as a repulsive gravitational source. Even though this component can be understood as a cosmological constant, or a vacuum energy density, the value needed for this constant differs greatly from what would be expected in quantum physics. Quantitatively, this dark energy component constitutes 70% of the total energy density, and therefore we only have a complete description for 5% of the total amount of energy in the Universe. Hence, unless one believes that the solution to the problem of dark matter is just waiting around the corner and gladly accepts the fine-tuned value of the cosmological constant, we are still far

from a closed cosmological model.

In this thesis we will focus on the problem of dark energy. The measurement of the accelerated expansion of the Universe from the dimming of type Ia supernovae [5, 6] (recently awarded with a Nobel prize for physics) was a surprising result. This discovery has spawned numerous theoretical models that try to describe and understand the nature of dark energy. Is it a cosmological constant? Is it due to a dynamical field? Do we have to modify the theory of gravity on large scales? Are we correctly interpreting the observations? If due to a vacuum energy density or some mechanism with similar dynamics, the effects of dark energy must be most evident at late times, since it must eventually dominate over the other components. Thus, studying the nature of dark energy is a very relevant motivation to explore the late-time evolution of the Universe. Galaxy redshift surveys are one of the most important sources of empirical data in this regime. Through them we can extract information from different observational probes, such as weak gravitational lensing, the study of clusters of galaxies or large-scale structure observables related to galaxy clustering. Our understanding of the recent history of the Universe has improved through previous and present galaxy surveys such as 2dfGRS or SDSS, and will continue to do so with forthcoming surveys, such as the Dark Energy Survey, BigBOSS, LSST or Euclid.

In sum, a wealth of high quality data coming from different observational sources is available to us. However, when confronted with large amounts of precise data, which must be used to constrain many different theoretical models, several concerns must be raised: first of all, for such large datasets an efficient processing of the data is indispensable. Also, statistical and systematic errors in the data, as well as the common systematics of different probes, must be thoroughly understood and correctly treated. Furthermore, in order to obtain unbiased constraints on various cosmological models, one must understand and minimize the assumptions made about the underlying model in the treatment of the data. Finally, many different observational effects must be correctly described for each model, and theoretical as well as astrophysical uncertainties must be accounted for. Each of these concerns poses a challenge which must be addressed in order to fully extract all the constraining power from the data.

After the review chapters 1 and 2, this thesis will touch on two topics related to the late-time evolution of the Universe and the problem of dark energy:

- **The baryon acoustic oscillations (BAO) as a standard ruler.** Pressure gradients in the baryon-photon plasma before photon decoupling create sound waves which leave an imprint in the late-time distribution of matter in the Universe, in the form of a characteristic length scale. The measurement of this scale provides a very robust determination of the distance-redshift relation, and can be used to improve our understanding of the expansion of the Universe at late times. This thesis covers the work described in the references [1, 2], in which a model-independent method to measure the BAO scale in the angular and radial directions from galaxy surveys was presented. This is discussed in chapter 3.

- **Void models as an alternative to dark energy.** In order to make sure that the observational data gives hints of new physics, we must verify the validity of all the assumptions made when interpreting that data. One key assumption of the standard cosmological model is the so-called Cosmological Principle: the Universe must look roughly the same as seen by every observer. The validity of this principle is not easy to prove, since we can not travel across cosmological distances to verify that the Universe is statistically the same everywhere. For this reason, some groups have explored the consequences and possibilities of dropping the Cosmological Principle. One possibility is that we might be living close to the centre of a very large underdense region (a void),

a clearly distinct location in the Universe. It has been noted that in this situation, observations interpreted in terms of a homogeneous cosmology, could mimic an accelerated expansion of the Universe. Our contribution to the study of void models [7, 8] has been to devise a technique to perform large-scale structure simulations of these models, and to use these simulations to understand them further. This is covered in chapter 4.

A summary of the main results obtained during the course of my Ph.D. studentship is presented in chapter 5. Appendices A, B and C contain some useful results and calculations that were used during this work.

CHAPTER 1

The standard model of cosmology

The standard model of cosmology is solidly founded on the basic and elegant Cosmological Principle (CP):

> "On sufficiently large scales, the Universe is homogeneous and isotropic."

Once this premise is accepted, the evolution of the Universe can be completely determined in terms of a few cosmological parameters that can be measured from astronomical observations.

In 1929 E. Hubble [9] observed that galaxies seemed to be moving away from us with a recessional velocity that is proportional to their separation distance (the so-called *Hubble Law*). This result, together with the CP, implies that earlier on these objects were not only closer to us, but to any other point in the Universe. Thus, the distance between any two points is growing: space itself is expanding. By extrapolating this process we must conclude that at some point in the past all the matter and energy in the Universe must have been in a very concentrated state. This led to the formulation of the *Big Bang Theory*, which, together with the theory of *inflation* [10, 11], constitute the cornerstones of the standard model of cosmology. This model successfully predicts the aforementioned expansion of the Universe, the relative abundances of light elements, the existence of the Cosmic Microwave Background (CMB), its spectrum of anisotropies (both in temperature and polarization) and the distribution of matter on large scales (large-scale structure (LSS)). In the first part of this chapter we will outline the mathematical framework behind this model and review the main events in the history of the Universe.

Years later, in 1998, the observation of the dimming of the light coming from type I-a supernovae (SNeIa) yielded an unexpected result [6, 5]: the recessional velocity of the SNe was growing. In a homogeneous universe this implies an accelerated expansion rate, which intuitively contradicts the result one would expect from a theory in which gravity is an ever-attractive force. The usual way to explain this is to introduce an unknown form of energy density (hence the name dark energy (DE)) that acts as a repulsive gravitational source. This component can be described as a cosmological constant or, more generally, as a fluid with negative pressure. Although this solves the problem of the accelerated expansion, the presence of this unknown energy poses many other problems. We will describe these problems and some alternative models of DE in the last part of this chapter.

1.1 The FRW metric

The way to interpret the CP in the framework of General Relativity (GR) is to say that our space-time is a 4-dimensional manifold in which the spatial sections (the hypersurfaces $t = \text{cte.}$, where t is a time-like coordinate, for example the comoving time) are maximally symmetric 3-dimensional manifolds, which are completely determined by their curvature scalar. The most general metric satisfying these conditions can be written, in comoving coordinates, as [12]

$$ds^2 = -dt^2 + a(t)^2 \left[\frac{dr^2}{1 - k\, r^2} + r^2 (d\theta^2 + \sin^2\theta\, d\phi^2) \right]. \tag{1.1}$$

This is the so-called Friedmann-Robertson-Walker (FRW) metric [13]. Notice that the distance between any two points measured at time t_1 will be a factor $a(t_1)/a(t_2)$ larger than measured at $t_2 < t_1$. For this reason the function $a(t)$ is called the *scale factor* and its dependence with t describes the expansion of the Universe. The constant k is proportional to the constant curvature scalar of the spatial hypersurfaces. Thus open, closed and flat universes are described by $k < 0$, $k > 0$ and $k = 0$ respectively. Since we can always redefine the units in which we measure distances, we can define the present value of the scale factor a_0 at will (the subindex $_0$ will label the present values of quantities from here on unless otherwise stated). Throughout this work we will adopt the commonly used gauge choice $a_0 = 1$.

1.1.1 Cosmological parameters

For simplicity, we will assume the energy content of the Universe, given by the energy-momentum tensor $T_{\mu\nu}$, to be made up of different non-interacting components, $T^{\mu\nu} = \sum_i T_i^{\mu\nu}$. Furthermore we will model each of these components as a perfect fluid with an equation of state, relating pressure p and density ρ, of the form $p_i = w_i\, \rho_i$ (notice that a cosmological constant would fit this definition with $w_\Lambda = -1$). Therefore $T_i^{\mu}{}_{\nu} = \rho_i\, \mathrm{diag}(1, -w_i, -w_i, -w_i)$. Since these components are decoupled from one another, energy-momentum conservation ($D_\mu T_\nu^\mu = 0$) holds separately for each of them. In our coordinates this is

$$\dot{\rho}_i + 3H(p_i + \rho_i) = 0, \tag{1.2}$$

where $H \equiv \frac{\dot{a}}{a}$. This is readily integrated for our simple equation of state:

$$\rho_i(t) \propto a^{-3(1+w_i)}. \tag{1.3}$$

Three types of components are usually considered: non-relativistic matter, or *dust* ($w_\mathrm{M} = 0$), relativistic matter, or *radiation* ($w_\mathrm{R} = 1/3$), and dark energy. As we will see later on, the standard model of cosmology describes dark energy as a cosmological constant Λ with $w_\mathrm{DE} = -1$, but other models predict different values for w_DE. Such is the importance of this parameter that one usually writes $w \equiv w_\mathrm{DE}$ (and we will do so from here on). As for dust, usually two distinct types of non-relativistic components are considered: baryonic matter, the ordinary matter we observe in galaxies and stars, and cold dark matter (CDM), a pressureless species which, unlike baryons, has a negligible interaction with the radiation (hence the prefix "dark"). Both of these components contribute with a^{-3} to H^2, but their different interplay with the radiation is crucial to understand many processes in the early Universe, as well as the spectrum of matter fluctuations we observe today. Multiple observations support the existence of CDM, which makes up about 80% of the total matter content of the Universe. The review [14] contains a wealth of information regarding our current knowledge of dark matter.

Now, inserting this energy-momentum tensor and the FRW metric into Einstein's equations we obtain Friedmann's equations:

$$H^2 = \frac{8\pi G}{3}\sum_i \rho_i - \frac{k}{a^2}, \tag{1.4}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\sum_i (\rho_i + 3p_i), \tag{1.5}$$

where we have defined $\rho_\Lambda \equiv \frac{\Lambda}{8\pi G}$.

It is now useful to define the following quantities:

$$\Omega_i(t) = \frac{8\pi G}{3H^2(t)}\rho_i(t), \quad \Omega_k(t) = \frac{-k}{H^2(t)\,a^2(t)}. \tag{1.6}$$

These are the so-called *cosmological parameters* and can be interpreted as the ratio of the density of some component to the critical density

$$\Omega_i = \frac{\rho_i}{\rho_c}, \tag{1.7}$$

where ρ_c is the density in a flat ($k = 0$) cosmology. In terms of these parameters the first Friedmann equation can be written in a nice closed form

$$H^2(t) = H_0^2\left[\sum_i \Omega_{i,0}\, a(t)^{-3(1+w_i)} + \Omega_{k,0}\, a(t)^{-2}\right], \tag{1.8}$$

$$\Omega_k(t) + \sum_i \Omega_i(t) = 1 \tag{1.9}$$

The first equation shows that knowing the present values of the cosmological parameters $\Omega_{i,0}$, $\Omega_{k,0}$ (we will drop the subindex $_0$ for the present cosmological parameters from here on and we will write an explicit time dependence when using them at any other time) together with the present Hubble parameter $H_0 \equiv 100\,h\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ and the initial condition $a(t_0) = 1$ is enough to completely determine the dynamics of $a(t)$. The second equation is the so-called *cosmic sum rule*.

The second Friedmann equation (1.5) is now trivial, since to write (1.8) we have made use of the energy-momentum conservation equation (1.2) and only two of the three are independent. However it is interesting to rewrite (1.5) in the following form:

$$q(t) \equiv -\frac{\ddot{a}a}{\dot{a}^2} = \frac{1}{2}\sum_i \Omega_i(t)(1 + 3\,w_i), \tag{1.10}$$

where the sum does not include Ω_k. Here $q(t)$ is the so-called *deceleration parameter*. Since an accelerating expansion occurs if $q < 0$, a fluid component will contribute to a positive acceleration as long as $w < -\frac{1}{3}$, which is the case for the cosmological constant. We will come back to this point in the next section.

To finish, it is worth rewriting eq. (1.8) in the presence of the components we have previously introduced (matter, radiation and vacuum energy)

$$H^2 = H_0^2\left[\Omega_{\mathrm{M}}\, a^{-3} + \Omega_{\mathrm{R}}\, a^{-4} + \Omega_{\mathrm{DE}}\, a^{-3(1+w)} + \Omega_k\, a^{-2}\right]. \tag{1.11}$$

1.1.2 Redshift and distances

Studying the light coming from luminous distant objects one realizes that the emission spectral lines are systematically displaced towards larger wavelengths. This *redshift* is due to an increase in the photon wavelength caused by the cosmic expansion that takes place during its journey to us. The redshift is defined as

$$z = \frac{\lambda_0 - \lambda}{\lambda}, \tag{1.12}$$

where λ_0 and λ are the emitted and received wavelengths, and can be related to the position of the emitter as well as to the time at which the photon was emitted. To do so we must first solve the geodesic equation for radial photons. Since we will derive it later in the more general case of LTB models (section 4.1.2), we will not dwell in the details and will just write the result:

$$a^2(t)\,\frac{d\chi}{d\tau} = \text{const.}, \quad \frac{d\,t}{d\tau} = a(t)\frac{d\chi}{d\tau}, \tag{1.13}$$

where χ is defined in eq. (1.17). Thus, the energy of the photon is

$$\Rightarrow \quad p^0 \propto \frac{1}{a(t)}, \tag{1.14}$$

and the ratio of energies is the inverse ratio of scale factors. Therefore it is easy to see that

$$a = \frac{1}{1+z}. \tag{1.15}$$

In order to use astronomical data to constrain the cosmological model, we will be interested in calculating distances between objects. However the only direct observables we have are angular positions in the sky and spectra. One can define different distances based on the different physical processes used to estimate them from the direct observables. These are some of the most important definitions [12, 15, 16]:

- **Comoving radial distance.** The photons emitted at t_E from a galaxy at r_g and reaching us at t_0, $r = 0$ travel along null geodesics (throughout this discussion we will place the observer at $r = 0$ for simplicity, but these results will not depend on this particular choice, since the FRW metric is homogeneous and isotropic). Hence

$$ds^2 = 0 \quad \Rightarrow \quad \int_0^{r_g} \frac{dr}{\sqrt{1-k\,r^2}} = \int_{t_E}^{t_0} \frac{dt}{a(t)}. \tag{1.16}$$

  Writing it in terms of redshift,

$$\chi(z) \equiv \int_0^{r(z)} \frac{dr'}{\sqrt{1-k\,r'^2}} = \int_0^{z} \frac{dz'}{H(z')}. \tag{1.17}$$

- **Proper physical radial distance.** Consider a set of comoving observers lying on a straight line between us ($r = 0$) and some galaxy at r_g, θ_g, φ_g, each observer being infinitesimally close to his neighbours. At a fixed cosmic time t all the observers measure the distance to the next one by, say, measuring the travel time of light signals. Later on we add all these measurements up to compute the total distance between us and the galaxy. This is the *proper distance*. The distance measured by each observer is

$$ds^2 = 0 \quad \Rightarrow \quad d(\text{distance}) = dt = a(t)\frac{dr}{\sqrt{1-k\,r^2}}. \tag{1.18}$$

Summing all such contributions yields

$$d_p(\chi, t) = a(t)\,\chi. \tag{1.19}$$

In practice it is impossible to set up such an experiment to measure the distance to a distant galaxy, therefore we have to define other distances, based on the practical means we have to make observations.

- **Luminosity distance.** In a Euclidean space, if L is the absolute luminosity of a point source at a distance d from an observer at O, the flux measured at O would be $F = L/(4\pi\, d^2)$. Therefore, given a source whose absolute luminosity L is known, we define the *luminosity distance* as

$$d_L = \sqrt{\frac{L}{4\pi\, F}}, \tag{1.20}$$

where F is the measured flux. Now imagine a point source at χ, θ, φ (notice that relative to this source we will also have a radial coordinate χ). The source emits photons at t_E and are observed by us at t_0. These photons are spread isotropically over a sphere formed by all the points with comoving radial distance χ. Looking at the FRW line element we can see that the proper area of this sphere is $A = 4\pi\, a_0^2\, r^2$. Also, the energies of these photons, as well as their arrival rate, are redshifted by a factor $1 + z$ (that makes 2 of these factors). With all this, the observed flux will be

$$F = \frac{L(t_E)}{4\pi(a_0\, r\,(1+z))^2}, \tag{1.21}$$

and we can see that the luminosity distance is:

$$d_L(z) = r(z)\,(1+z) \tag{1.22}$$

- **Angular diameter distance.** Consider, in a Euclidean space, two objects separated by a distance D. As observed from a point at a distance d from both of them, they would subtend an angle $\Delta\theta = \frac{D}{d}$ (where we have assumed $d \gg D$). Therefore, for an object of proper length l subtending an angle $\Delta\theta$ we define the *angular diameter distance* by

$$d_A = \frac{l}{\Delta\theta}. \tag{1.23}$$

Now, for two objects sitting at r, θ, φ and r, $\theta + \Delta\theta$, φ (since the space is isotropic we can always choose our coordinates such that any two points have the same φ coordinate) the proper distance between them at some time t_E (at which they emit two lightrays towards us) is $l = a(t_E)\, r\, \Delta\theta$, therefore

$$d_A(z) = \frac{r(z)}{1+z} = \frac{d_L(z)}{(1+z)^2}. \tag{1.24}$$

It is interesting to note that the last equality must hold in any metric theory [17].

1.1.3 Brief history of the Universe

The key to understanding the thermal history of the Universe is comparing the interaction rates of the species in the cosmic fluid with the rate of expansion [18, 19]. From this comparison one can deduce when a given species decouples from the rest and ceases to contribute to the total thermal degrees of freedom of the plasma. We can estimate the interaction rate as [18]

$$\Gamma_{\rm int} = \langle n\,\sigma\,v\rangle, \tag{1.25}$$

where n is the number density of target particles, σ is the cross section of the interaction, v is the relative velocities of the particles involved and $\langle\,\rangle$ indicates averaging over a thermal distribution. It can be proven that when $\Gamma_{\rm int} < H$ the species interact less than once on average. At this point we say that the particles are decoupled.

As we go back in time the Universe gets denser and warmer (in fact one can prove that $T \propto a^{-1}$), thus at earlier times there is more energy available for the interactions. According to most theories of inflation an inflationary epoch starts around the Planck era ($10^{19}\,$GeV, $10^{-43}\,$s). At the end of inflation the energy density of the inflaton is transformed into relativistic particles in a process called *reheating*. There are many different reheating models, and no direct experimental evidence to test most of them, however the situation at the end of this process is always the same: the Universe is dominated by relativistic species. When the Universe cools down further the baryons that have been generated (mainly quarks) form protons and neutrons ($10^{2}\,$MeV, $10^{-5}\,$s).

The furthest window into the past history of the Universe that we have direct evidence of is the stage of *nucleosynthesis* (1 MeV, 1 s), when the temperature is low enough to allow protons and neutrons to form bound systems, creating the first nuclei (deuterium, helium, lithium). At about the same time neutrinos decouple and electrons and positrons annihilate generating large amounts of photons.

As the Universe cools down the number of relativistic degrees of freedom decreases, eventually being overrun by the non-relativistic ones: matter-radiation equality occurs (1 eV, $10^5\,$yrs). Soon afterwards the temperature allows electrons to form bound systems with the nuclei forming the first atoms in a process called *recombination* (0.3 eV, 380000 yrs). Later the photons cease to interact with these electrons and the Universe "becomes transparent". These photons are allowed to travel freely and form what we observe today as the Cosmic Microwave Background.

Before photon decoupling baryons and radiation were tightly coupled in a single imperfect fluid. Pressure forces in this fluid prevent gravitational collapse in the baryonic component which only falls in the potential wells set up by dark matter once it decouples. During matter domination perturbations in the matter component grow more rapidly and the first structures appear. In the current standard model, large-scale structure formation occurs in a "bottom-up" picture: small structures form first and then merge and accrete more material to form larger ones. The first galaxies form at $t \sim 0.5\,$Gyrs and afterwards merge forming groups and clusters of galaxies, the most massive gravitationally bound objects ever observed, with masses up to $10^{15}\ M_\odot/h$.

Since the density of non-relativistic matter falls like a^{-3}, it will eventually become very diluted and in the presence of a cosmological constant ($\rho_\Lambda = $ const.) it will be overpowered by it. Thus at late times the Universe becomes Λ-dominated.

1.2 Late-time acceleration and the problem of dark energy

Since the first observation of the dimming of hight redshift type Ia supernovae by Riess et al. and Perlmutter et al. [6, 5], an ever-increasing amount of evidence has appeared that supports the conclusion that the late expansion of the Universe is accelerated. Here are the most important observations:

- **Age of the Universe.** The age of the Universe can be calculated by integrating Friedmann's equation (1.8). Assuming that only ordinary non-relativistic matter exists in a flat universe, this is $t_{BB} = \frac{2}{3}H_0^{-1}$. With the measured value of the local expansion rate $h = 0.72 \pm 0.08$ the age of the Universe in this case would be 8.2 Gyrs $< t_{BB} <$ 10.2 Gyrs. However the ages of some globular clusters have been estimated using different methods to be $\gtrsim 13$ Gyrs [20]. A way to circumvent this inconsistency is to introduce a non-zero cosmological constant or a negative spatial curvature. Since the curvature $|\Omega_k|$ has been tightly constrained by CMB observations to be much smaller than unity, it is impossible to solve this problem without some form of dark energy (at least in an FRW cosmology). With the currently accepted values of the cosmological parameters ($\Omega_{\rm M} \simeq 0.3$,$\Omega_\Lambda \simeq 0.7$, $h \simeq 0.7$, [21]) the age of the Universe $t_{BB} \simeq 13.5$ Gyrs, well beyond the age of globular clusters.

- **SNeIa.** Supernova explosions can be classified according to the absorption lines of chemical elements. Type II supernovae have a hydrogen absorption line (Type I do not). Type Ia have an ionized silicon line, Type Ib have a helium line and Type Ic lack both. The absolute luminosity of SNeIa is well understood: it is almost constant at the peak of brightness and can be estimated through their light curves [22]. Thus Type Ia SNe can be used as standard candles to measure the distance-redshift relation $d_L(z)$ by measuring their relative magnitudes (fluxes). As we said, it was the observation of hight-redshift supernovae by Riess et al. [6] that gave the first hint of a non-zero dark energy component and since then further SNeIa data have supported this conclusion (e.g. [23, 24]).

- **Baryon acoustic oscillations (BAO).** Before photon decoupling baryonic matter and radiation are tightly coupled and can be described as a single imperfect fluid with non-zero pressure. Pressure gradients generate sound waves in the fluid and prevent gravitational collapse until baryons cease to feel the photon drag. Afterwards the baryonic component behaves just like dark matter. This process produces a small "bump" (the so-called BAO peak) in the 2-point correlation function (2PCF) of the distribution of galaxies at the comoving scale of the sound horizon at the drag epoch r_s (see section 2.2 for more details). Thus r_s can be used as a standard ruler to probe the distance-redshift relation by measuring the position of the BAO peak at different redshifts. This has been done by different groups [25, 26, 27, 28, 29, 30] (see figure 1.1 for a summary of the most relevant measurements), and these measurements are consistent with the current values of the cosmological parameters [4].

- **CMB.** The positions and amplitudes of the acoustic peaks in the angular power spectrum of temperature anisotropies in the CMB can also be used to constrain the values of the cosmological parameters (in fact they provide the tightest bounds). A full and thorough account of CMB physics lies outside the scope of this work and we refer the reader to [31] for more information. Suffice it to say that CMB measurements imply that the Universe must be very close to flat and are consistent with the results above in every other respect [32, 21].

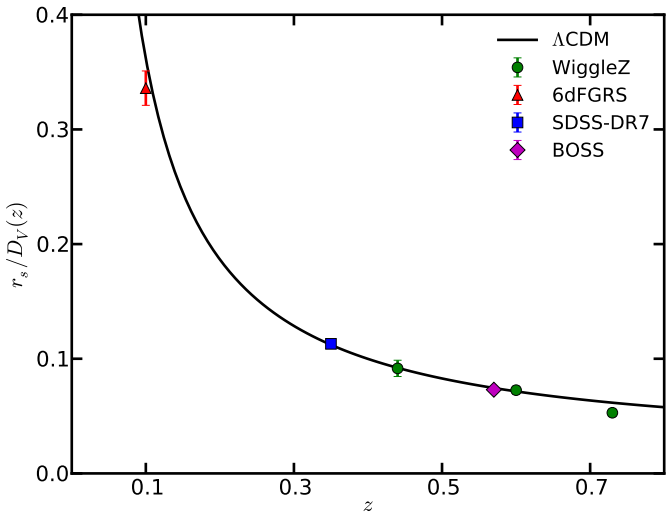

Figure 1.1: Most relevant BAO measurements in terms of r_s/D_V, estimated from the position of the BAO peak in the monopole of the 2PCF. Data taken from [25, 26, 27, 28, 30].

When faced with such overwhelming evidence one can only accept the accelerated expansion of the Universe driven by an otherwise unobserved form of energy (or at least that the distance-redshift relation is not that of a homogeneous universe with ordinary energy components). A myriad of different approaches to this problem have been proposed, let us review some of them.

1.2.1 Cosmological constant

The standard and simplest approach to the problem of dark energy is to attribute its effects to a non-zero cosmological constant. Effectively a cosmological constant acts as a constant energy density (independent of the expansion of the Universe) and is therefore interpreted as the energy density of the vacuum. This idea is not new [33] and indeed such vacuum energy is predicted by Quantum Field Theory (QFT), however it has a couple of serious theoretical problems. Namely, the energy density of the vacuum is estimated in QFT as

$$\rho_{\rm vac} = \frac{c}{2\pi^2\hbar^3}\int_0^{E_c} E(p)\, p^2\, dp \sim \frac{E_c^4}{8\pi^2(\hbar\, c)^3}, \tag{1.26}$$

where E_c is the ultraviolet cut-off of the theory, beyond which it is not believed to be valid. If we take this cut-off to be at the Planck scale, $E_c \sim M_P \simeq 10^{19}\,\mathrm{GeV}/c^2$ (GR is believed to be valid up to this scale) we obtain $\rho_{\rm vac} \sim 10^{95}$ kg m^{-3}, which is 121 orders of magnitude larger than the measured value $\rho_\Lambda \sim 10^{-26}$ kg m^{-3} (although see [34] for a more thorough calculation). Even using the electroweak scale as a cut-off we would still find a disagreement of 65 orders of magnitude. One may think that there exists a symmetry in nature that exactly cancels these ultraviolet divergences, or that the zero point energy is just an artefact of the field ordering in the theory, but this would not yield a *very* small vacuum energy density, it would be exactly 0, and therefore the fine-tuning problem remains.

Another issue is the so-called *coincidence* problem. When we calculate the time of matter-Λ equality or the moment at which the expansion begins to accelerate we realize that both

events take place remarkably close to the present time:

$$z_{\rm eq} = \left(\frac{\Omega_\Lambda}{\Omega_{\rm M}}\right)^{\frac{1}{3}} - 1 \simeq 0.3,$$
$$z_{\rm acc} = \left(\frac{2\,\Omega_\Lambda}{\Omega_{\rm M}}\right)^{\frac{1}{3}} - 1 \simeq 0.7. \tag{1.27}$$

There is no reason why this should be so, and this coincidence is viewed as a problem of the ΛCDM model. In this framework these two problems are obviously related. Other models of dark energy can get rid of the first one, but the fact that the positive acceleration must start in the near past is necessary in all cases to allow the growth of structure.

A way to "explain" the low value of the vacuum energy density is to use anthropic arguments: consider the possibility that $\rho_{\rm vac}$ could take a number of different values (a *very* large number according to string theory [35]). Then the value it takes in our Universe must be compatible with our existence, since we are here to observe it. Using this kind of arguments in 1987 Weinberg [36] gave bounds to the possible values of the cosmological constant, within which the measured quantity lies. Of course, the impossibility to probe the values of the vacuum energy density in other universes puts anthropic arguments beyond any hope of experimental proof.

1.2.2 Alternative dark energy models

As stated in the previous section, one could think that there exists some symmetry in nature that exactly cancels the energy density of the vacuum. In this case there must be some exotic component with negative pressure driving the accelerated expansion of the Universe. Since there exist many different alternative models of dark energy and we will not be interested in them for the rest of this work, we will only describe here the most popular ones briefly. For further information we refer the reader to [37, 38, 39].

- **Quintessence.** Like most models of inflation, quintessence models use a canonical scalar field as a source of accelerated expansion [40]. Unlike a cosmological constant, the energy density of quintessence does not need to be very small with respect to that of radiation and matter in the early Universe, and therefore the problem of the fine tuning of Λ can be alleviated. In these models (as in most alternative models) the equation of state of DE varies with time, and therefore deviations from $w \equiv -1$ are searched for. A popular parametrization of this variation is [41, 42]

$$w = w_0 + w_a(1-a), \tag{1.28}$$

  and in this case the dark energy density varies as

$$p_{\rm DE} = (w_0 + w_a(1-a))\rho_{\rm DE} \quad \Rightarrow \quad \rho_{\rm DE}(a) \propto a^{3(1+w_0+w_a)}\,\exp\left(3\,w_a\,(1-a)\right).$$

- ***k*-essence.** In quintessence models the cosmic acceleration is caused by a canonical scalar field with a slowly varying potential. *k*-essence models, in turn, exploit the possibility of realizing an accelerated expansion through a non-canonical kinetic term [43].

- **Chameleons.** The existence of strongly coupled scalar fields is not in general compatible with local gravity experiments. An attempt at solving this problem is the chameleon mechanism, by which the mass of the scalar is made to depend on the environmental density [44].

- **Vector dark energy.** Before the recent measurements made by CMS and ATLAS regarding what could be the Higgs boson [45, 46], no fundamental scalar fields had been observed. For this reason some models have been developed in which the accelerated expansion is driven by the energy density of a vector field [47].

- **Generalized Chaplygin gas.** In these models the DE fluid has an unusual equation of state:
$$p \propto -\rho^{-\alpha}. \tag{1.29}$$
It has been shown that, at least at the background level, this could account for both dark matter and dark energy [48].

- Other models include phantoms, coupled dark energy, scaling solutions, etc. More details can be found in the references above.

1.2.3 Modified gravity models

There exist many observational constraints and many local experiments that show that GR describes the gravitational interaction correctly up to scales of the size of the Solar System (even galactic scales). However there is no empirical proof of this validity on cosmological scales and we may wonder whether the accelerated expansion of the Universe could be due to corrections to GR on these scales without the need for a dark component (although the aforementioned experiments severely restrict such modifications - [49]).

The simplest and most popular models are the so-called "$f(R)$" models, in which the gravitational action is given by some general function of the Ricci curvature scalar R:

$$S = \frac{1}{8\pi G}\int d^4x\,\sqrt{-g}f(R) \,+\, S_{\text{matter}}. \tag{1.30}$$

Even in the early 80s it was known that a quadratic correction $f(R) = R + \alpha R^2$ [50] could be responsible for inflation in the early Universe, since the quadratic term gave rise to asymptotically exact de Sitter solutions. It is possible to explain an accelerated expansion in the past with this mechanism, but it is of no use for our purposes.

One can also consider, not only corrections to the action, but also the possibility that the connection $\Gamma^{\mu}_{\nu\sigma}$ may not be just the Christoffel symbols, but an independent field with respect to which the action must also be varied (this is the so-called Palatini formalism) [51].

Other viable modifications of gravity include: *scalar-tensor* theories, which use a scalar field that is non-trivially coupled to gravity, *Gauss-Bonnet* DE models, in which the modified action also depends on other contractions of the curvature tensors ($R_{\mu\nu}$ and $R_{\mu\nu\rho\sigma}$), and *Braneworld* theories, motivated by string theory, in which standard model particles are confined on a 3D brane embedded in a 5-dimensional bulk spacetime with large extra dimensions. For more information on modified gravity models see [52].

1.2.4 Accelerated expansion without DE

It has been argued ([53]) that all the observations listed at the beginning of this section measure the distance-redshift relation, but do not give any direct evidence of the existence of an exotic form of energy driving an accelerated expansion of the Universe. In an FRW (that is, spatially homogeneous) universe the distance-redshift relation is given by eqs. (1.8), (1.22) and (1.24), and under this assumption the observations certainly point towards a non-zero dark energy component. However we should first make sure that these equations are applicable, that is, that the assumption of homogeneity on large scales is correct, before drawing any

conclusions. There exist two different approaches that try to explain the measured distance-redshift relation using the inhomogeneous distribution of ordinary matter:

- **Backreaction.** The reasoning behind the use of the FRW metric to describe the Universe is that, since the distribution of matter on large scales is homogeneous, one should use a spatially homogeneous metric to describe it. However, the mathematically correct process would be to use the exact metric to calculate Einstein's tensor and then average Einstein's equations over large scales. Since Einstein's equations are non-linear both methods are not equivalent. Unfortunately it is impossible to know the form of the exact metric a priori and we can only try to quantify the error that is made by using the first method. The extra terms due to this backreaction of perturbations can indeed be estimated and could in principle contribute to an apparent acceleration of the expansion. One could argue that since the perturbations in the Newtonian potential are small ($\Phi \sim 10^{-5}$, see section 2), these contributions cannot be large enough to account for the observed acceleration. However, even if Φ is small, density perturbations are proportional to second derivatives of Φ (Poisson's equation) and need not be small (and in fact are not). The bad news is that this effect is very poorly understood and, furthermore, since it is purely general-relativistic, it cannot be tested by Newtonian simulations. See [54, 55, 56, 57, 58] for a more detailed description and [59] for a first attempt at comparing SNeIa data with backreaction.

- **Void models.** A more aggressive approach is to consider the possibility that we might live inside a large underdense region. The observations of the distribution of matter around us and the CMB imply that this underdensity (void) must be spherically symmetric and that we must live close to its centre. Since astronomical observations are performed in the lightcone (that is, we see the objects as they were at the time the observed photons were emitted), any variation with redshift is interpreted within a homogeneous scheme as a variation in time, while it would be a combination of temporal and spatial variations in an inhomogeneous cosmology:

$$\text{Homogeneous} \rightarrow \frac{d}{dz} = \frac{dt}{dz}\frac{d}{dt},$$
$$\text{Inhomogeneous} \rightarrow \frac{d}{dz} = \frac{dt}{dz}\frac{\partial}{\partial t} + \frac{dr}{dz}\frac{\partial}{\partial r}.$$

  Thus if we observe an expansion rate $H(z)$ at some $z > 0$ that is lower than the locally measured value, we would say that there has been an acceleration in an FRW framework, while this could be easily reproduced without acceleration by a void model, in which the rate of expansion decreases as the distance to the centre of the void grows. These models were proposed soon after the first supernovae data supporting cosmic acceleration [60, 61] and their study has experienced a recent rebirth. See [62, 63, 64, 65, 66] for more details. Of course such a strong violation of the Copernican Principle is very unattractive from both an aesthetic and a philosophical point of view. However, as with everything else in science, no philosophical prejudice can rule out a theoretical model, only empirical observations can.

  In the chapter 4 we will describe the mathematical framework of void models, their performance when put to the test of different cosmological observations, as well as our contribution to their study.

CHAPTER 2

Cosmological structure formation

Even though the standard cosmological model is based on the premise that the Universe is homogeneous and isotropic on large scales, it is evident to anyone who has ever looked at the sky at night that it is not so on smaller scales. Structures with very different sizes can be found, from stars and planets at the AU scale (~ 4 μpc) to galaxies at the kpc scale and clusters and superclusters of galaxies at the Mpc scale. As we understand them, these structures have their origin in small linear perturbations at early times in an otherwise homogeneous Universe, which have grown to reach their present non-linear state via gravitational collapse. The theory of inflation gives a fairly satisfactory explanation of the origin of these perturbations in terms of quantum fluctuations in the primordial density field, while the theory of cosmological perturbations and large scale structure studies and models their evolution throughout the history of the Universe.

The description of the origin of the cosmological perturbations within the inflationary paradigm is a fascinating and beautiful result, and its generic predictions can be tested through the available probes of the early Universe, such as the temperature and polarization anisotropies in the CMB. A full account of the theory of inflation lies well outside the scope of this work (although the references [19, 67, 68, 11] make for a very enjoyable read), however it will be useful to state its most relevant generic results:

- Cosmological perturbations have their origin in quantum fluctuations in the primordial energy density field, which grow at later times via gravitational collapse to form the non-linear structures we observe today.

- Most models of inflation predict that these perturbations can be described as a Gaussian (or very close to Gaussian) random field.

- The primordial spectrum of perturbations is very close to scale-invariance. Deviations from scale-invariance are usually parametrized in terms of the scalar spectral index n_s:

$$P_\Phi(k) \propto k^{n_s-1}, \tag{2.1}$$

  where k is the wavenumber in Fourier space and $P_\Phi(k) \propto \langle|\Phi_{\mathbf{k}}|^2\rangle$.

After the end of inflation and reheating, relativistic species dominate the energy density of the Universe. In this primordial plasma, baryons are tightly coupled to photons and both evolve as a single fluid, while CDM is decoupled from them. A full study of the evolution

of perturbations at this stage demands a fully relativistic approach to perturbation theory, the details of which go beyond our focus. The main results of this treatment are, however, extremely important to understand the spectrum of late-time perturbations, and we will describe them in section 2.2.

The usual first approach to structure formation is to model the matter and energy content of the Universe as a slightly perturbed homogeneous and isotropic perfect fluid and to treat it in linearized Newtonian theory, which is the topic of the next section. Even though this model gives useful results in certain regimes, it has some clear shortcomings. First of all, a Newtonian treatment is not able to capture many relativistic effects, like the evolution of perturbations on scales close to or larger than the horizon or the behaviour of perturbations in the relativistic species. As we said above, a covariant and gauge-invariant framework is needed for this description. More importantly for structure formation purposes, due to the nature of gravitational collapse, inhomogeneities in the matter density grow, reaching non-linear values sooner or later. Although higher orders in perturbation theory (PT) can help in understanding the growth of structure in mildly non-linear scales, the perturbative series eventually breaks down. At this point one must make use of either numerical methods or phenomenological models to study structure formation. We will give a description of these approaches in sections 2.3 and 2.4.

2.1 Newtonian linear perturbation theory

The basic equations for a perfect fluid in Newtonian theory are

$$\partial_t \rho + \nabla \cdot (\rho\, \mathbf{V}) = 0, \tag{2.2}$$

$$\partial_t \mathbf{V} + (\mathbf{V} \cdot \nabla)\mathbf{V} + \frac{\nabla p}{\rho} + \nabla \phi = 0, \tag{2.3}$$

$$\partial_t s + (\mathbf{V} \cdot \nabla)\, s = 0, \tag{2.4}$$

$$\nabla^2 \phi = 4\pi G \rho, \tag{2.5}$$

$$p = p(\rho, s), \tag{2.6}$$

where (2.2) is the continuity equation, (2.3) is Euler's equation, (2.4) comes from entropy conservation, (2.5) is Poisson's equation for the Newtonian gravitational potential and (2.6) is the equation of state of the fluid.

When applied to a perturbed homogeneous fluid in an expanding background, one finds the following equations to 0th and 1st order in the perturbations:

- 0th order:

$$\dot{\rho_0} + 3\rho_0 H = 0, \quad \dot{H} + H^2 = -\frac{4\pi G}{3}\rho_0. \tag{2.7}$$

  These two can be derived from Friedmann's equations (1.5).

- 1st order:

$$\dot{\delta} + \frac{1}{a}\nabla \cdot \mathbf{v} = 0 \tag{2.8}$$

$$\dot{\mathbf{v}} + H\,\mathbf{v} + \frac{c_s^2}{a}\nabla\,\delta + \frac{1}{a}\nabla\Phi = 0 \tag{2.9}$$

$$\nabla^2 \Phi = 4\pi G\, a^2 \rho_0\, \delta. \tag{2.10}$$

Here c_s^2 is the sound speed in the fluid, and we have expanded the density field $\rho(t,\mathbf{q}) = \bar{\rho}(t)\,(1+\delta(t,\mathbf{q}))$, the velocity field $\mathbf{V}(t,\mathbf{q}) = \dot{a}\,\mathbf{q}+\mathbf{v}(t,\mathbf{q})$, the gravitational potential $\Phi_T(t,\mathbf{q}) = \Phi_0(t,\mathbf{q})+\Phi(t,\mathbf{q})$ and the pressure $p(t,\mathbf{q}) = \bar{p}(t)+c_s^2\,\bar{\rho}(t)\,\delta(t,\mathbf{q})$ around their background values. Several approximations must be considered before reaching this result, in particular we have neglected entropy perturbations. Note that here all spatial derivatives are taken with respect to comoving coordinates $\mathbf{q}$, related to physical coordinates by $\mathbf{x} = a(t)\,\mathbf{q}$.

2.1.1 Density perturbations

Combining (2.8) with the divergence of (2.9) and then with (2.10) yields a second-order equation for δ alone:

$$\ddot{\delta}_\mathbf{k} + 2H\,\dot{\delta}_\mathbf{k} + \frac{c_s^2}{a^2}\,(k^2 - k_J^2(t))\,\delta_\mathbf{k} = 0, \tag{2.11}$$

where we have rewritten everything in terms of the Fourier transform $\delta_\mathbf{k}$ defined as in equation (A.2), and we have defined the Jeans wavenumber $k_J^2(a) \equiv 4\pi G\,a^2\,\rho_0/c_s^2$ [69]. This scale separates two types of behaviours:

1. **Small scales:** For $k \gg k_J$ the solution in the WKB approximation corresponds to plane waves with an amplitude that decays as $\delta \propto a^{-1/2}$.

2. **Large scales:** For $k \ll k_J$ the general solution is

$$\delta_\mathbf{k}(t) = C_\mathbf{k}^{(1)} H(t) + C_\mathbf{k}^{(2)} H(t) \int^t \frac{dt'}{H^2(t')\,a^2(t')}, \quad (k \ll k_J). \tag{2.12}$$

It is easy to see that if we switched off the expansion ($H = 0$), the growing-mode solution would correspond to an exponential growth $\delta \propto \exp[\sqrt{4\pi G\bar{\rho}}\,t]$. However, during matter domination ($a \propto t^{2/3}$) we have a growing mode $\delta \propto a \propto t^{2/3}$. The background expansion decelerates the growth of structure, and this is good news, since otherwise the perturbative approach would soon cease to yield any reliable results. Another important thing in the large-scale regime is that the evolution equation (2.11) does not depend on k (since spatial derivatives are suppressed). Thus the only scale-dependence is in the initial conditions: perturbations evolve in a self-similar way, preserving their initial shape.

Some important conclusions can be drawn from these solutions:

1. On small scales, below the sound horizon, pressure forces dominate over gravity, so that perturbations behave like sound waves with a decaying amplitude ($\propto a^{-1/2}$).

2. On large scales the matter perturbations can grow, but their growth is decelerated by the background expansion.

3. On large scales (and on any scale in the absence of pressure, which is the case for perturbations in a decoupled dust-like component, such as CDM) inhomogeneities grow in a self-similar way, preserving their initial shape.

Normally one only considers the growing mode for the growth of perturbations, given by the second solution in eq. (2.12) (this is only a good approximation for adiabatic perturbations [70]). This is a function of time alone known as the *growth factor* and is usually labelled $D(a)$. Since at early times, during matter domination, the growth factor is proportional to the scale factor, it is customary to normalize it to $D(a \ll 1) = a$. With this normalization the growth factor is

$$D(a) = \frac{5}{2}\,\Omega_\mathrm{M}\,H_0^2\,H(a) \int_0^a \frac{da'}{[a'\,H(a')]^3}. \tag{2.13}$$

At late times, in the presence of matter and vacuum energy, this integral can be approximated by [71]

$$D(a) = \left[(\Omega_{\rm M}(a))^{4/7} - \Omega_\Lambda(a) + \left(1 + \frac{\Omega_{\rm M}(a)}{2}\right)\left(1 + \frac{\Omega_\Lambda(a)}{70}\right)\right]^{-1}. \tag{2.14}$$

Particular solutions

The perturbation equation (2.11) can be solved analytically in the large-scale limit in a flat background filled with dust and another unperturbed component with an equation of state w. The solution can be found by changing the variable in the differential equation (2.11) $t \to a(t)$ and then transforming it into the hypergeometric equation, yielding [72]

$$\begin{aligned}\delta = & C_1\, a \cdot {}_2F_1\left(\frac{w-1}{2w}, -\frac{1}{3w}, \frac{6w-5}{6w}; \frac{\Omega_{\rm M}-1}{\Omega_{\rm M}} a^{-3w}\right) + \\ & C_2\, a^{-3/2} \cdot {}_2F_1\left(\frac{3w+2}{6w}, \frac{1}{2w}, \frac{6w+5}{6w}; \frac{\Omega_{\rm M}-1}{\Omega_{\rm M}} a^{-3w}\right),\end{aligned} \tag{2.15}$$

where ${}_2F_1(a,b,c;z)$ is the Gauss hypergeometric function. In particular for a universe filled with dust and radiation ($w = 1/3$) the solution can be expressed in terms of elementary functions:

$$\delta = A_1\left(x + \frac{2}{3}\right) + A_2\left[\left(x + \frac{2}{3}\ln\left(\frac{\sqrt{1+x}+1}{\sqrt{1+x}-1} - 2\sqrt{1+x}\right)\right)\right], \tag{2.16}$$

with $x \equiv a/a_{\rm eq}$ ($a_{\rm eq}$ being the scale factor at matter-radiation equality). This is the so-called *Meszaros solution* [73], and will be important when we try to determine the shape of the CDM power-spectrum. The Meszaros solution shows the effect of the background expansion on the growth of perturbations: during radiation domination ($x \ll 1$) the expansion is faster and the growth is logarithmic with a at most (matter perturbations are almost frozen). During matter domination ($x \gg 1$) the expansion slows down and the growing mode $\propto a$ is recovered. Thus we see that the relativistic species affect the growth of matter perturbations by changing the rate of expansion.

For a flat universe with a cosmological constant ($w = -1$) the growing mode is

$$D(a) = a \cdot {}_2F_1\left(1, \frac{1}{3}, \frac{11}{6}; -\frac{\Omega_\Lambda}{\Omega_{\rm M}} a^3\right) \tag{2.17}$$

2.1.2 The velocity field

Let us separate the Fourier modes of the peculiar velocity field into longitudinal and transverse (irrotational and divergence-free in real space) components

$$v_{\mathbf{k},\parallel} \equiv \mathbf{v}_\mathbf{k} \cdot \frac{\mathbf{k}}{k}, \quad \mathbf{v}_{\mathbf{k},\perp} \equiv \mathbf{v}_\mathbf{k} - v_{\mathbf{k},\parallel}\frac{\mathbf{k}}{k}. \tag{2.18}$$

Euler's equation (2.9) for the transverse component is

$$\dot{\mathbf{v}}_{\mathbf{k},\perp} + H\,\mathbf{v}_{\mathbf{k},\perp} = \frac{1}{a}\frac{d(a\,\mathbf{v}_{\mathbf{k},\perp})}{dt} = 0 \quad \Rightarrow \quad \mathbf{v}_{\mathbf{k},\perp} \propto \frac{1}{a}. \tag{2.19}$$

Therefore the rotational component of the peculiar velocity vanishes rapidly and one can consider $\mathbf{v}$ as a purely longitudinal vector $\mathbf{v}_\mathbf{k} \simeq v_\mathbf{k}\frac{\mathbf{k}}{k}$. The continuity equation for $v_\mathbf{k}$ relates it linearly to the density contrast as

$$v_\mathbf{k} = -i\frac{a\,H}{k}\, f\, \delta_\mathbf{k}, \tag{2.20}$$

where we have defined the growth rate

$$f \equiv \frac{d\ln(\delta_{\mathbf{k}})}{d\ln(a)}. \tag{2.21}$$

Notice that, in principle, the growth factor is a function of both $\mathbf{k}$ and t (or a). However, we will see that at late times density perturbations grow self-similarly:

$$\delta_{\mathbf{k}} \propto T_k \times D(a), \tag{2.22}$$

where T_k is a k-dependent *transfer function*, and therefore all scale dependence of the growth rate f cancels. Even though there exist exact solutions for $f(a)$ in wCDM universes [74, 72], it has become customary to describe the growth factor in terms of the growth index γ defined as

$$f(a) = \Omega_{\rm M}(a)^{\gamma} \tag{2.23}$$

Within GR and for realistic values of the cosmological parameters, the growth index is $\gamma \simeq 0.55$.

In section 2.1.3 we will be interested in relating the velocity field to the galaxy number overdensity. Assuming that galaxies are fair tracers of the underlying matter density, the galaxy and matter overdensities should be the same. However, as we will see, we now know that galaxies are in fact biased with respect to the matter distribution, and most models of galaxy formation predict that galaxies should form around density peaks [75]. Modelling this biasing mechanism is a complicated task (see section 2.4, [76]), however, a linear bias factor $\delta_g \simeq b\,\delta_{\rm M}$ seems to be a good approximation on large scales. In this approximation the galaxy overdensity is related to the velocity field via the continuity equation

$$v_{\mathbf{k}} = -i\frac{a\,H}{k}\,\beta\,\delta_{g,\mathbf{k}}, \tag{2.24}$$

where we have defined the factor $\beta \equiv f/b$.

Finally, let us note that in equation (2.20) the wavenumber k appears in the denominator, showing the sensitivity of peculiar velocities to large (rather than small) scale structure. This makes observational probes related to peculiar velocities, such as redshift-space distortions (RSD) [77], very important for LSS.

2.1.3 Redshift-space distortions

For most discrete probes of the matter density field (e.g. galaxies), their radial position can be estimated through their redshift, which is measured, for example, from the shift in their emission spectra. However, due to the presence of density perturbations, galaxies have non-zero radial peculiar velocities, which alter their measured redshift through an additional Doppler effect. This distortion is correlated with the real density perturbation, and therefore changes the statistics of the observed matter distribution. This effect was first modelled by N. Kaiser in the famous paper [78] and has come to be widely known as the *Kaiser effect*. We will not give a thorough derivation of all the results displayed in this section, and the reader is referred to the very useful review by A. Hamilton [77] for further details.

The Kaiser effect

Assuming peculiar velocities are small, the observed redshift of an object z_o is altered from its comoving redshift z (the shift due only to the expansion) by $z_o = z + v_z$, where v_z is the

projection of the object's peculiar velocity along the line of sight (in units of c). Thus, this object will be assigned a radial comoving distance given by

$$\chi_s \equiv \chi(z + v_z) \simeq \chi_r + \frac{v_z}{H(z)}, \tag{2.25}$$

where we have used equation (1.17). From this point on, all quantities measured in redshift space will be labelled by a subscript s, while real-space observables will be subscripted with r. Since the total number of galaxies must be conserved in the coordinate transformation $\chi_r \to \chi_s$, it is easy to estimate the observed number overdensity in terms of the real-space overdensity and velocity fields

$$\delta_s = \delta_r - \left(\frac{\alpha(\chi)}{\chi} + a\partial_t + \frac{1}{\chi^2}\partial_\chi \chi^2 \right) \frac{v_z}{H}, \tag{2.26}$$

where α is the logarithmic slope of the real-space mean number density $\alpha \equiv d\log \bar{n}_r / d\log \chi$, and we have linearized for small v_z.

Our aim is to derive the effect of RSDs on the two-point statistics of the galaxy number density. In the plane-parallel approximation (i.e. only for pairs of galaxies whose relative separation r_{12} is much smaller than distance to the observer $r_{12} \ll \chi$), it is possible to see that all the operators scaling with $1/\chi$ inside the brackets in eq. (2.26) are subdominant, and we are left with a simplified relation

$$\delta_s = \delta_r - \frac{\partial_\chi v_z}{H(z)}. \tag{2.27}$$

Assuming that vector modes in the velocity field vanish (i.e. the velocity can be written as a pure gradient), it is possible to use the continuity equation (2.24) to relate the real and redshift-space overdensities at the linear level:

$$\delta_s = \left(1 + \beta \partial_z^2 \nabla^{-2}\right) \delta_r, \tag{2.28}$$

where ∂_z denotes differentiation along the line of sight and we have defined the integral operator

$$\nabla^{-2}\delta(\mathbf{x}) \equiv -\frac{1}{(2\pi)^{3/2}} \int \frac{dk^3}{k^2} e^{i\,\mathbf{k}\cdot\mathbf{x}}\, \delta_{\mathbf{k}}. \tag{2.29}$$

This relation takes a much simpler form in Fourier space

$$\delta_{s,\mathbf{k}} = (1 + \beta\, \mu_{\mathbf{k}}^2)\, \delta_{r,\mathbf{k}}, \tag{2.30}$$

where $\mu_{\mathbf{k}} \equiv \mathbf{k}\cdot\hat{\mathbf{n}}_z/k$ is the cosine of the angle between the wave vector $\mathbf{k}$ and the line of sight $\hat{\mathbf{n}}_z$. Hence, redshift-space distortions induce an anisotropy in the observed power spectrum, which now is a function of both k and $\mu_{\mathbf{k}}$

$$P_s(k, \mu_{\mathbf{k}}, z) = (1 + \beta(z)\, \mu_{\mathbf{k}}^2)^2\, P_r(k, z). \tag{2.31}$$

Fingers of god

It is possible to understand qualitatively the effect of redshift-space distortions on the observed matter density both in the linear and non-linear regimes.

Consider a spherical overdensity. All the masses will tend to fall towards the centre of the clump, and the infalling velocities will be higher for smaller distances. While velocities in the directions transverse to the line of sight ($\mathrm{LOS}_\perp$) do not modify the observed redshift,

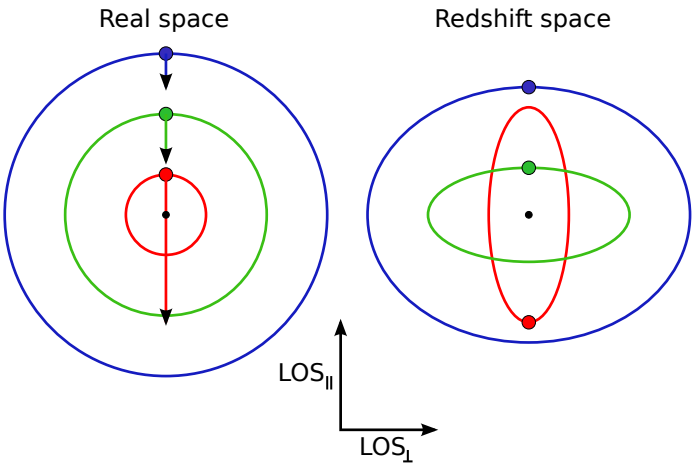

Figure 2.1: Illustrative representation of RSDs. Small radial velocities, far away from density peaks, cause a squashing of the observed structure due to the misplacement of the galaxies along the line of sight. This effect is enhanced as we get closer to the centre of the cluster, to the point that a turnaround point occurs, beyond which, the collapsing shells appear inside out in redshift space.

longitudinal ($\mathrm{LOS}_{\parallel}$) velocities do. On large (linear) scales, this is only a mild effect, and spherical distributions in real space look melon-shaped in redshift space. However, on small (non-linear) scales these velocities are large enough to compensate Hubble's law and the spheres are turned inside out and deformed in shape, forming structures that point towards the observers, commonly known as fingers of god (FoG, see figure 2.1).

While it is easy to describe the effect of RSDs in the linear and plane-parallel approximation, the behaviour in the non-linear regime and the FoG are more difficult to model. There exist several prescriptions in the literature for this modelling [79, 80, 81, 82], an example of which is the streaming model of [81]

$$P_s(k,\mu_{\mathbf{k}},z) = (1+\beta(z)\,\mu_{\mathbf{k}}^2)^2\,F(k,\mu_{\mathbf{k}})\,P_r(k,z), \qquad F(k,\mu_{\mathbf{k}}) = \frac{1}{(1+k^2\,\mu_{\mathbf{k}}^2\,\Sigma_v^2)^2}. \tag{2.32}$$

Measuring RSDs

While at first sight RSDs seem to be a nuisance, an observational effect that we would rather not have, they turn out to be a very powerful observational tool. Redshift-space distortions allow us to obtain information about the interplay between the density and velocity fields, and thus can be used as probes of the nature of gravitational collapse and the growth of structure.

Since it is impossible to determine the overall normalization of the two-point correlation function using only two-point galaxy clustering data, due to the galaxy bias, it is not possible to constrain the growth rate f alone, but rather the combination $f\,\sigma_8$. This has been done for several datasets [83, 84, 85]. In most recent cases, a likelihood analysis is performed on the correlation function multipoles, which we describe here.

As we have seen, in redshift space the power spectrum depends on the projection of $\mathbf{k}$ along the line of sight, and therefore it can be expanded in terms of Legendre polynomials of

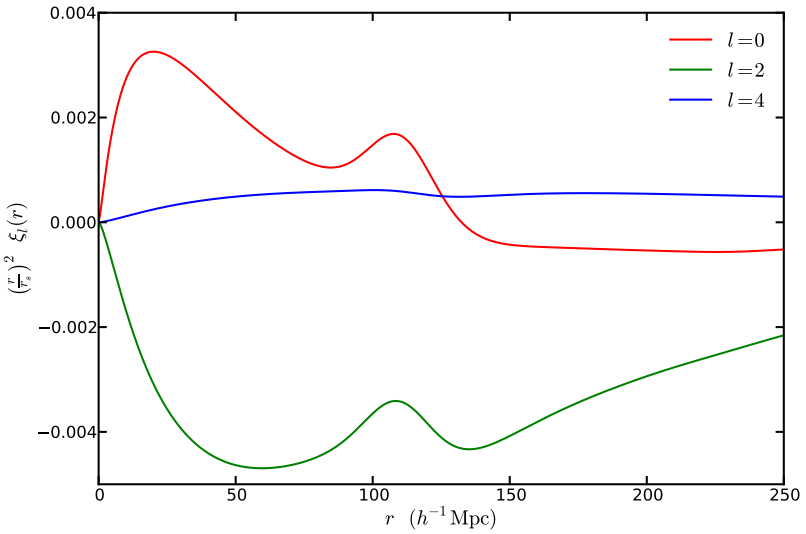

Figure 2.2: First three non-zero multipoles of the two-point correlation function in redshift space as predicted in the Kaiser approximation.

$\mu_{\mathbf{k}}$:

$$P(k,\mu_{\mathbf{k}},z) = \sum_{l=0}^{\infty} P_l(k,z)\,L_l(\mu_{\mathbf{k}}) \longrightarrow P_l(k,z) \equiv \frac{(2l+1)}{2}\int_{-1}^{1} P(k,\mu_{\mathbf{k}},z)\,L_l(\mu_{\mathbf{k}})\,d\mu_{\mathbf{k}}, \quad (2.33)$$

where the functions $P_l(k,z)$ are the multipoles of the power spectrum. A similar decomposition can be performed for the correlation function, which now is a function of both r and $\mu \equiv \mathbf{r}\cdot\hat{\mathbf{n}}_z/r$, and the multipoles of the correlation function can be calculated in terms of the P_l's:

$$\xi(r,\mu) = \sum_{l=0}^{\infty} \xi_l(r)\,L_l(\mu), \quad \xi_l(r) = \frac{i^l}{2\pi^2}\int_0^{\infty} k^2\,j_l(kr)\,P_l(k)\,dk, \quad (2.34)$$

where j_l is the l-th spherical Bessel function. Further details about the relation between the anisotropic power spectrum and the configuration-space correlation function can be found in section A.1.5.

Notice that, since the power spectrum is an even function of $\mu_{\mathbf{k}}$, only the odd multipoles vanish. Furthermore, in the Kaiser approximation, all multipoles with $l \geq 4$ are exactly 0. Thus most analyses of RSD are performed on the first three multipoles: the monopole ($l = 0$), quadrupole ($l = 2$) and hexadecupole ($l = 4$). We have plotted these three multipoles for a WMAP7-compatible cosmology in figure 2.2, and their explicit expressions in terms of the real-space function can be found in section A.1.5.

2.2 From the drag epoch to the late-time power spectrum

The aim of this section is to describe the different physical processes that shape the distribution of matter perturbations we observe in cosmological surveys. Many of the features of the spectrum of inhomogeneities can be understood by studying the evolution of perturbations in the baryon-photon fluid before photon decoupling. The main results of this study are detailed in section 2.2.1. The shape of the late-time matter power spectrum can then be understood

in terms of the distribution of matter perturbations after baryons cease to feel the radiation drag (the drag epoch), since they grow self-similarly afterwards. This is described in section 2.2.2.

The presence of a fraction of baryonic matter in the total contribution of dust-like species induces important effects in the shape of the power spectrum which can potentially be used as observational probes. These are discussed in section 2.2.3.

Finally, radial peculiar velocities alter the observed redshift of objects and hence distort the spectrum of perturbations. As discussed in section 2.1.3, this distortion enables us to test the physics of structure formation and the growth of inhomogeneities.

2.2.1 Physics of perturbations before the drag epoch

As has been previously described, in order to understand the evolution of inhomogeneities in the early Universe, a fully covariant and gauge-invariant treatment of metric perturbations is needed. Although this problem is now well understood, a thorough description would be too involved for our purposes, and only the main results will be presented here. The reader is referred to [19, 67, 86, 87] for further details.

Before recombination and the drag epoch, baryons are tightly coupled to photons, and the baryon-radiation component can be treated as a single imperfect fluid. The other relevant species is cold dark matter, which only interacts with the baryon-photon plasma through gravity. Thus energy and momentum are conserved separately for both fluids. Throughout the rest of this section we will label baryon, photons and CDM by b, γ and c respectively.

The $b-\gamma$ fluid has a non-zero viscosity, which must be included in the perturbation equations. This enables photons to transfer energy between different regions of the fluid through diffusion, effectively dissipating perturbations on scales smaller than the photon mean free path.

The full set of linearized equations, assuming adiabatic initial conditions, is [19]

$$(a(\delta_c - 3\Phi)')' - a\,\nabla^2\Phi = 0, \quad \delta_\mathrm{b} = \frac{3}{4}\delta_\gamma \tag{2.35}$$

$$\left(\frac{\delta_\gamma'}{c_s^2}\right)' - \frac{3\kappa}{a\,\rho_{\gamma,0}}\nabla^2\delta_\gamma' - \nabla^2\delta_\gamma = \frac{4}{3\,c_s^2}\nabla^2\Phi + \left(\frac{4\Phi'}{c_s^2}\right)' - \frac{12\,\kappa}{a\,\rho_\gamma}\nabla^2\Phi', \tag{2.36}$$

$$\nabla^2\Phi - 3\mathcal{H}\,(\Phi' + \mathcal{H}\Phi) = 4\pi\,G\,a^2\left(\rho_c\,\delta_c + \frac{1}{3\,c_s^2}\rho_\gamma\,\delta_\gamma\right), \tag{2.37}$$

where the prime ($'$) denotes differentiation with respect to the conformal time $\eta \equiv \int a(t)\,dt$, κ is the shear viscosity coefficient, related to the photon mean free path τ_γ by $\kappa = \frac{4}{15}\rho_\gamma\tau_\gamma$, and c_s is the speed of sound in the fluid:

$$c_s^2 = \frac{1}{3}\left(1 + \frac{3}{4}\frac{\rho_\mathrm{b}}{\rho_\gamma}\right)^{-1}. \tag{2.38}$$

Without further approximations these equations can only be solved numerically. There exist several public Boltzmann codes, such as `CAMB` [88] or `CLASS` [89], which can solve these very fast for a broad set of cosmological models, and they are widely used by the cosmology community to compare the predictions of linear perturbation theory with observations of the CMB and LSS. This approach, however scientifically useful, does not provide a descriptive view of the general behaviour of the different components. For this reason we would like to describe here the main features of these solutions.

Growth before and after equality

As can be seen from the Meszaros solution (2.16), the growth of matter perturbations is radically different before and after matter-radiation equality:

- $t < t_{\rm eq}$. During radiation domination, the growth of matter perturbations on subhorizon scales is suppressed, growing logarithmically with a at best. At the same time, the non-zero pressure gradients generate sound waves in the radiation perturbations.
- $t > t_{\rm eq}$. After equality matter perturbations grow faster ($D(a) \sim a$) and in a self-similar way. The Newtonian potential stays constant on all scales.

Horizon effects

A relativistic treatment of perturbation theory introduces a characteristic length scale: the scale of the horizon. This scale separates causally connected and disconnected events, and thus defines the size of regions over which gravitational collapse can be efficient. A general result from gauge-invariant perturbation theory for adiabatic fluctuations is the fact that metric perturbations on scales larger than the horizon do not evolve:

$$\Phi_{\mathbf{k}}(\eta) = \text{constant}, \quad k\eta \ll 1. \tag{2.39}$$

After inflation, all scales of interest are outside the horizon, and, therefore, perturbations entering the horizon at different epochs will have very different histories.

Different histories

Due to horizon effects, a good way to understand the evolution of matter perturbations is to study their properties before and after entering the horizon at the two most relevant events: matter-radiation equality and photon decoupling.

- $k \ll \eta_{\rm drag}$. For perturbations which have remained outside the horizon until after the drag epoch, the only evolution is a drop in the amplitude of metric perturbations by a factor 9/10.
- $\eta_{\rm drag} \ll k \ll \eta_{\rm eq}$. In this case perturbations enter the horizon well inside matter domination but before photon decoupling. The dark matter evolves decoupled from the the baryon-photon plasma, and its amplitude grows with the scale factor. The Newtonian potential does not change except for the already mentioned 9/10 factor. On the other hand, the baryon-photon fluid behaves differently: pressure gradients generate sound waves with a phase $k\, r_s(\eta)$, where r_s is the comoving sound horizon

  $$r_s(t) = \int_0^t \frac{c_s(t)}{a(t)} dt, \tag{2.40}$$

  which defines the largest distance a sound wave may have travelled since the Big-Bang. The amplitude of these oscillations is damped on scales below the photon-diffusion scale due to the non-zero viscosity of the fluid. This effect is commonly called *Silk damping* [90], and appears in the perturbation equations as a Gaussian damping term $\propto \exp[-(k/k_D)^2]$, where the damping scale k_D is

  $$k_D^2 = \left(\frac{2}{5} \int c_s^2 \, \frac{\tau_\gamma}{a} \, d\eta \right)^{-1} \tag{2.41}$$

- $k \gg \eta_{\rm eq}$. In this case matter perturbations enter the horizon during radiation domination, and their growth at early times is suppressed according to the Meszaros solution.

2.2.2 Transfer functions and the CDM model

Early on, at the end of inflation and the beginning of the radiation epoch, all interesting scales are outside the horizon. In this regime the potential is constant and all modes behave in the same fashion. Later on two events occur: a given scale enters the horizon and the Universe becomes matter-dominated (for simplicity we will forget about baryons and recombination for the moment). Since scales entering the horizon before and after equality have very different histories, we must expect a feature of some sort in the power spectrum of matter perturbations at the scale of the horizon at equality $k_{\rm eq} \sim \eta_{\rm eq}^{-1}$. At late times all the interesting modes are inside the horizon again and evolve self-similarly. We would like to relate the perturbations we observe at late times to the primordial spectrum set up by inflation. Let us define the *transfer function* as [87]

$$\Phi_{\mathbf{k}}(a) = \frac{9}{10}\Phi_{0,\mathbf{k}} T_k \frac{D(a)}{a}, \tag{2.42}$$

Here the transfer function T_k links the amplitude of the primordial perturbation $\Phi_{0,\mathbf{k}}$ with the amplitude of the perturbation at some time $a_{\rm late} \gg a_{\rm eq}$

$$T_k = \frac{\Phi_{\mathbf{k}}(a_{\rm late})}{\frac{9}{10}\Phi_{0,\mathbf{k}}} \tag{2.43}$$

(the extra factor 9/10 should now be familiar and is only extracted for convenience). That T_k does not depend on $a_{\rm late}$ or on the direction of $\mathbf{k}$ is granted by the self similar evolution at late times and by the fact that the evolution equations depend only on $k = |\mathbf{k}|$ (at least at first order in PT). The growth factor $D(a)$ (see eq. (2.13)) relates the amplitude of the perturbation at this late time $a_{\rm late}$ with its value at any later time a:

$$\frac{D(a)}{a} = \frac{\Phi_{\mathbf{k}}(a)}{\Phi_{\mathbf{k}}(a_{\rm late})}, \quad (a \gg a_{\rm eq}). \tag{2.44}$$

Assuming a Newtonian treatment is valid (after all at late times we mainly care about sub-horizon perturbations in the non-relativistic component), we can write a similar expression for the matter perturbation $\delta_{\mathbf{k}}$ using Poisson's equation (2.10) in Fourier space:

$$\delta_{\mathbf{k}}(a) = -\frac{k^2\, a}{\frac{3}{2}\Omega_{\rm M}\, H_0^2}\Phi_{\mathbf{k}}(a) = \frac{3}{5}\frac{k^2}{\Omega_{\rm M}\, H_0^2}\Phi_{0,\mathbf{k}} T_k\, D(a). \tag{2.45}$$

Using the inflationary prediction for the primordial spectrum of perturbations (eq. (2.1)), the matter power spectrum can be written in terms of the transfer function:

$$P(k,z) = 2\pi^2 \delta_H^2 \left(\frac{k}{H_0}\right)^{n_s} \frac{1}{H_0^3} T_k^2\, D^2(z). \tag{2.46}$$

Here the amplitude and spectral tilt of the primordial power spectrum are parametrized by δ_H and n_s. Remember that most inflationary scenarios predict an almost scale-invariant spectral tilt ($n_s \sim 1$), which is supported by CMB observations ($n_s \simeq 0.96$, [4]). See appendix A for a precise definition of the power spectrum $P(k) \propto \langle|\delta_{\mathbf{k}}|^2\rangle$ and other 2-point statistics of random fields. A more common way to parametrize the amplitude of matter perturbations is through the σ_8 parameter (instead of δ_H), which is defined as the variance of the linear matter density contrast in spheres with a radius of $8\,{\rm Mpc}\,h^{-1}$ at redshift 0. This can be calculated as

$$\sigma_R^2(z) \equiv \frac{1}{2\pi^2}\int_0^\infty k^2\, |W(kR)|^2\, P(k,z)\, dk, \quad W(x) \equiv 3\,\frac{\sin x - x\,\cos x}{x^3}. \tag{2.47}$$

As we have seen, the form of the matter power spectrum is determined by three scales:

- The scale of the horizon at equality, $k_{\rm eq}^{-1}$. This separates modes that entered the horizon during the radiation and matter eras and correspondingly the two asymptotic behaviours of the power-spectrum:

$$T_k \propto \begin{cases} 1, & k \ll k_{\rm eq} \\ k^{-2}\ln(k), & k \gg k_{\rm eq} \end{cases} \tag{2.48}$$

- The sound horizon at decoupling, r_s. Defined as:

$$r_s = \int_0^{\eta_{\rm dec}} c_s\, d\eta = \frac{2}{3\,k_{\rm eq}}\sqrt{\frac{6}{R_{\rm eq}}}\ln\left(\frac{\sqrt{1+R_{\rm dec}}+\sqrt{R_{\rm dec}+R_{\rm eq}}}{1+\sqrt{R_{\rm eq}}}\right), \tag{2.49}$$

  with $R = \frac{3}{4}\frac{\rho_{\rm b}}{\rho_\gamma}$. This is the maximum distance a sound wave could have travelled before photon decoupling. Oscillations in the baryon-photon fluid prior to recombination leave an imprint on the matter power spectrum with phase $k\,r_s$ and a peak in the correlation function at $r \sim r_s$.

- The damping scale, k_D. The oscillatory terms are damped by a factor $e^{-(k/k_D)^2}$, caused by the viscosity in the imperfect baryon-photon fluid.

For most practical cases we have $k_D^{-1} < r_s < k_{\rm eq}^{-1}$.

2.2.3 Fitting functions and baryonic features

Analytic solutions for the matter perturbations can be found in some asymptotic scales (e.g. well outside or inside the horizon, at early or late times, etc.), however, very interesting signatures lie in the missing intermediate wavenumbers, like the bending point at $k = k_{\rm eq}$. For this reason numerical methods must be used in order to obtain precise quantitative results. Beside the already mentioned Boltzmann codes, several fitting functions have been developed that can reproduce the transfer functions with a reasonable accuracy for a wide range of cosmological models. A very simple parametrization was originally given in the famous paper by Bardeen, Bond, Kaiser and Szalay [75] (BBKS hereon):

$$T_k = \frac{\ln(1+2.34\,q)}{2.34\,q}\left[1+3.89\,q+(16.1\,q)^2+(5.46\,q)^3+(6.71\,q)^4\right]^{-1/4}, \tag{2.50}$$

where $q \equiv k/(\Omega_{\rm M}\,h)$. A more precise and very popular parametrization was developed by Eisenstein & Hu [91] and has been widely used. One advantage of this parametrization is that it allows for an intuitive understanding of the different effects that act on the shape of the matter power spectrum. On cosmological scales there are two main effects due to the non-zero baryonic component:

- For small wavelength modes entering the horizon before the drag epoch, the perturbations in the baryonic matter density cannot grow, since they are tightly coupled to photons and perturbations in the radiation density are frozen on sub-horizon scales. Afterwards baryons decouple and can fall into (and contribute to) the gravitational potential wells. These have been set up mainly by the cold dark matter, which started to collapse effectively after matter-radiation equality. Hence we can predict that a non-negligible baryonic contribution will reduce the power on small scales.

- Acoustic oscillations in the baryon-photon fluid before decoupling leave an imprint on the matter power spectrum, in the form of a sinusoidal contribution, with a frequency

given by the value of the sound horizon at the drag epoch r_s and an amplitude proportional to the baryonic fraction. This can be observed in the real-space two-point correlation function (see appendix A and chapter 3) as an excess probability at a distance corresponding to r_s.

2.3 Non-linearities I: analytic and phenomenological approaches

Up until now we have dealt with the linearized version of the perturbation equations. The solutions to these can only fully describe the growth of structure for a limited range of times and scales. In order to obtain more accurate predictions for LSS in the mildly non-linear and non-linear regimes, several analytical and phenomenological approaches have been proposed in the literature. We will review some of them here. Analytic results can be obtained by including the non-linear terms of the perturbation equations in a perturbative series. The main results of this approach are presented in section 2.3.1, and those of the alternative analytic treatment of Lagrangian perturbation theory are discussed in section 2.3.2. Non-linearities can also be described in a phenomenological way within the so-called *halo model*, which is described in section 2.3.3.

2.3.1 Higher-order corrections

We will describe here the effects on structure formation of including of non-linear terms in perturbation theory. It must be noted that a perturbative approach to structure formation will only be valid as long as inhomogeneities remain in the perturbative regime ($\delta \leq 1$) and the series will not converge in general otherwise. Nevertheless important insight can be gained about mildly non-linear scales through these first corrections. These topics are reviewed in [92].

We will focus only on Newtonian perturbations in the absence of pressure. This is not a bad approximation for the late Universe and for scales larger than the typical sizes of clusters and virialized structures ($l \gtrsim 1\,\mathrm{Mpc}\,h^{-1}$).

The equations

The continuity, Euler and Poisson equations fully describe the evolution of the dark matter density field in an expanding background:

$$\dot{\delta} + \frac{1}{a}\nabla \cdot ((1+\delta)\mathbf{v}) = 0, \tag{2.51}$$

$$\dot{\mathbf{v}} + H\mathbf{v} + \frac{1}{a}(\mathbf{v}\cdot\nabla)\mathbf{v} + \frac{1}{a}\nabla\Phi = 0, \tag{2.52}$$

$$\nabla^2\Phi = \frac{3}{2}a^2H^2(a)\Omega_{\mathrm{M}}(a)\delta. \tag{2.53}$$

Assuming irrotational fluid motions and defining the divergence of the velocity field $\theta \equiv \nabla\cdot\mathbf{v}$, these equations take a simpler form in Fourier space

$$\delta'_{\mathbf{k}} - \theta_{\mathbf{k}} = -\int dk_1^3\, dk_2^3\, \delta_{12}\, \alpha(\mathbf{k}_1, \mathbf{k}_2)\, \delta_{\mathbf{k}_1}\theta_{\mathbf{k}_2}, \tag{2.54}$$

$$\theta'_{\mathbf{k}} + \mathcal{H}\theta_{\mathbf{k}} + \frac{3}{2}\mathcal{H}^2\Omega_{\mathrm{M}}(a)\delta_{\mathbf{k}} = -\int dk_1^3 dk_2^3\, \delta_{12}\, \beta(\mathbf{k}_1, \mathbf{k}_2)\, \theta_{\mathbf{k}_1}\theta_{\mathbf{k}_2}$$

where $\delta_{12} \equiv \delta^{\mathcal{D}}(\mathbf{k}-\mathbf{k}_1-\mathbf{k}_2)$ and we have defined the kernels

$$\alpha(\mathbf{k}_1,\mathbf{k}_2) \equiv \frac{\mathbf{k}_{12}\cdot\mathbf{k}_2}{k_2^2}, \quad \beta(\mathbf{k}_1,\mathbf{k}_2) \equiv \frac{k_{12}^2(\mathbf{k}_1\cdot\mathbf{k}_2)}{2k_1^2k_2^2}, \quad \mathbf{k}_{12}\equiv\mathbf{k}_1+\mathbf{k}_2. \tag{2.55}$$

As we did before, $'$ denotes differentiation with respect to the conformal time η, and $\mathcal{H}\equiv a'/a$.

We have already found the solution to the linearized version of these equations

$$\begin{aligned}\delta_{\mathbf{k}}(a) &= \delta^{(1)+}_{\mathbf{k},0} D_1^+(a) + \delta^{(1)-}_{\mathbf{k},0} D_1^-(a)\\ \theta_{\mathbf{k}}(a) &= -\mathcal{H}\left[\delta^{(1)+}_{\mathbf{k},0} f(a) D_1^+(a) + \delta^{(1)-}_{\mathbf{k},0} g(a) D_1^-(a)\right]\end{aligned}$$

where $D_1^{+,-}$ are the linear growing and decaying solutions (normalized to the EdS solutions at early times) and

$$f(a) \equiv \frac{d\,\ln D_1^+}{d\,\ln a}, \quad g(a) \equiv \frac{d\,\ln D_1^-}{d\,\ln a}. \tag{2.56}$$

Non-linear corrections

Defining $\theta_{\mathbf{k}}(a) \equiv -\mathcal{H}(a)\,f(a)\,\lambda_{\mathbf{k}}(a)$, the non-linear equations (2.54) can be recast as:

$$\frac{d\,\delta_{\mathbf{k}}}{d\,\ln D_1^+(a)} - \lambda_{\mathbf{k}} = \int dk_1^3 dk_2^3\,\delta_{12}\,\alpha(\mathbf{k_1},\mathbf{k_2})\,\lambda_{\mathbf{k}_1}\,\delta_{\mathbf{k}_2}, \tag{2.57}$$

$$\frac{d\,\lambda_{\mathbf{k}}}{d\,\ln D_1^+(a)} + \lambda_{\mathbf{k}}\left(\frac{3}{2}\frac{\Omega_{\rm M}(a)}{f^2(a)} - 1\right) - \frac{3}{2}\frac{\Omega_{\rm M}(a)}{f^2(a)}\,\delta_{\mathbf{k}} = \int dk_1^3 dk_2^3\,\delta_{12}\,\beta(\mathbf{k_1},\mathbf{k_2})\,\lambda_{\mathbf{k}_1}\,\lambda_{\mathbf{k}_2}. \tag{2.58}$$

For an Einstein-deSitter (EdS) background ($\Omega_{\rm M}(a)=1$, $D_1^+(a)=a$, $f(a)=1$) these equations simplify further

$$\begin{aligned}\frac{d\,\delta_{\mathbf{k}}}{d\,\ln a} - \lambda_{\mathbf{k}} &= \int dk_1^3 dk_2^3\,\delta_{12}\,\alpha(\mathbf{k_1},\mathbf{k_2})\,\lambda_{\mathbf{k}_1}\delta_{\mathbf{k}_2},\\ \frac{d\,\lambda_{\mathbf{k}}}{d\,\ln a} + \frac{1}{2}\lambda_{\mathbf{k}} - \frac{3}{2}\,\delta_{\mathbf{k}} &= \int dk_1^3 dk_2^3\,\delta_{12}\,\beta(\mathbf{k_1},\mathbf{k_2})\,\lambda_{\mathbf{k}_1}\,\lambda_{\mathbf{k}_2}.\end{aligned}$$

We know the linear solution to these: $\lambda^{(1)}_{\mathbf{k}}(a) = \delta^{(1)}_{\mathbf{k}}(a) = \delta^{(1)}_{0,\mathbf{k}}\,a$. The perturbative process would now be to write $\delta_{\mathbf{k}}(a) = \delta^{(1)}_{\mathbf{k}}(a) + \delta^{(2)}_{\mathbf{k}}(a)$, substitute on the equations above, linearize them and solve for $\delta^{(2)}$. Then we would do the same for a third-order $\delta^{(3)}$, etc... However it can be shown [93, 94] that the perturbative series can be wrapped up as:

$$\delta_{\mathbf{k}}(a) = \sum_{n=1}^{\infty} \delta^{(n)}_{0,\mathbf{k}}\,a^n\ ; \quad \theta_{\mathbf{k}}(a) = -\mathcal{H}(a)f(a)\sum_{n=1}^{\infty}\lambda^{(n)}_{0,\mathbf{k}}\,a^n, \tag{2.59}$$

with $\lambda^{(1)}_{0,\mathbf{k}} = \delta^{(1)}_{0,\mathbf{k}}$ and:

$$\delta^{(n)}_{0,\mathbf{k}} = \int\left(\prod_{i=1}^{n} dk_i^3\right)\delta\left(\mathbf{k}-\sum_i \mathbf{k}_i\right) F_n(\mathbf{k}_1,...,\mathbf{k}_n)\left(\prod_i \delta^{(1)}_{0,\mathbf{k}_i}\right) \tag{2.60}$$

$$\lambda^{(n)}_{0,\mathbf{k}} = \int\left(\prod_{i=1}^{n} dk_i^3\right)\delta\left(\mathbf{k}-\sum_i \mathbf{k}_i\right) G_n(\mathbf{k}_1,...,\mathbf{k}_n)\left(\prod_i \delta^{(1)}_{0,\mathbf{k}_i}\right). \tag{2.61}$$

The kernels F_n and G_n are given recursively by

$$F_n(\mathbf{k}_1,...,\mathbf{k}_n) = \sum_{m=1}^{n-1} \frac{G_m(\mathbf{k}_1,...,\mathbf{k}_m)}{(2n+3)(n+1)} \left[(2n+1)\,\alpha(\mathbf{q}_1,\mathbf{q}_2)\,F_{n-m}(\mathbf{k}_{m+1},...,\mathbf{k}_n) + \right.$$
$$\left. +2\,\beta(\mathbf{q}_1,\mathbf{q}_2)\,G_{n-m}(\mathbf{k}_{m+1},...,\mathbf{k}_n)\right],$$
$$G_n(\mathbf{k}_1,...,\mathbf{k}_n) = \sum_{m=1}^{n-1} \frac{G_m(\mathbf{k}_1,...,\mathbf{k}_m)}{(2n+3)(n+1)} \left[(3\,\alpha(\mathbf{q}_1,\mathbf{q}_2)\,F_{n-m}(\mathbf{k}_{m+1},...,\mathbf{k}_n) + \right.$$
$$\left. +2n\,\beta(\mathbf{q}_1,\mathbf{q}_2)\,G_{n-m}(\mathbf{k}_{m+1},...,\mathbf{k}_n)\right],$$

with $F_1 = G_1 = 1$, $\mathbf{q}_1 = \mathbf{k}_1 + ... + \mathbf{k}_m$, $\mathbf{q}_2 = \mathbf{k}_{m+1} + ... + \mathbf{k}_n$.

One can see that the equations for any general cosmology (2.57,2.58) are not homogeneous in a and there is, in principle, no easy way to find a recursion relation for the different orders of the perturbative series such as the one we have found for EdS. However, under the approximation $f(a) \simeq [\Omega_{\rm M}(a)]^{1/2}$ we recover the equations for the EdS case with the change $a \to D_1^+(a)$. Hence the perturbative series can be again wrapped up in the same fashion as before:

$$\delta_{\mathbf{k}}(a) = \sum_{n=1}^{\infty} \delta_{0,\mathbf{k}}^{(n)}\,[D_1^+(a)]^n\ ; \quad \theta_{\mathbf{k}}(a) = -\mathcal{H}(a)f(a)\sum_{n=1}^{\infty} \lambda_{0,\mathbf{k}}^{(n)}\,[D_1^+(a)]^n, \tag{2.62}$$

with the integration constants $(\delta_{0,\mathbf{k}}^{(n)}, \lambda_{0,\mathbf{k}}^{(n)})$ given recursively in terms of $\delta_{0,\mathbf{k}}^{(1)}$ through eqs. (2.60, 2.61). We have seen that in the standard cosmological model, $f(a) = [\Omega_{\rm M}(a)]^{\gamma}$ is a fairly precise parametrization for the growth rate with $\gamma \simeq 0.55$, very close to the value of $1/2$ needed for this approximation.

2.3.2 Lagrangian perturbation theory

In Lagrangian perturbation theory (LPT) the object of study is not the perturbations of the density field in "constant" coordinates, but the displacement in the original positions of the fluid elements that cause (and are originated by) those perturbations. There exist several review articles in the literature describing this approach in depth. See, for example [95, 96].

The displacement field

In the absence of perturbations, the physical positions $\mathbf{x}$ of comoving objects are related to their original comoving coordinates $\mathbf{q}$ via

$$\mathbf{x} = a(t)\,\mathbf{q},$$

and in this case the dynamics are described by the scale factor $a(t)$. In the presence of perturbations, however, the dynamics are inhomogeneous and an overall scale factor is not enough to relate $\mathbf{q}$ to $\mathbf{x}$. In general one can write

$$\mathbf{x} = a(r)\,\mathbf{r}(t,\mathbf{q}) = a(t)\,[\mathbf{q} + \mathbf{\Psi}(t,\mathbf{q})]\,, \tag{2.63}$$

which defines the displacement vector $\mathbf{\Psi}(t,\mathbf{q})$. We have defined the auxiliary $\mathbf{r}$, which is the outcome of subtracting the Hubble expansion from the physical coordinates.

Under the Newtonian approximation the movement of a test particle is described by Newton's second law, which in terms of $\mathbf{r}$ and conformal time reads

$$\mathbf{r}'' + \mathcal{H}\mathbf{r}' = -\nabla_r \Phi. \tag{2.64}$$

The Newtonian potential Φ can be related to the perturbations of the density field caused by the displacements through Poisson's equation, and in turn, δ can be related to $\mathbf{r}$ through the conservation of mass

$$\bar{\rho}\, dq^3 = \bar{\rho}\,(1+\delta)\, dr^3 \quad \Rightarrow \quad 1+\delta = (\det(\hat{J}))^{-1} \equiv J^{-1}, \tag{2.65}$$

where $\hat{J}$ is the Jacobian matrix of the $\mathbf{q} \to \mathbf{r}$ transformation:

$$\hat{J}_{ij} = \frac{\partial r_i}{\partial q_j} = \delta_{ij} + \partial_j \Psi_i. \tag{2.66}$$

Taking the divergence of (2.64) and using Poisson's equation we obtain:

$$\nabla_r(\boldsymbol{\Psi}'' + \mathcal{H}\boldsymbol{\Psi}') = \frac{3}{2}\mathcal{H}^2\,\Omega_{\rm M}\,\frac{J-1}{J}, \tag{2.67}$$

and since $\nabla_r = \hat{J}^{-1}\cdot\nabla_q$:

$$J\,(\hat{J}^{-1}\nabla)\cdot(\boldsymbol{\Psi}'' + \mathcal{H}\boldsymbol{\Psi}') = \frac{3}{2}\,\mathcal{H}^2\,\Omega_{\rm M}\,(J-1). \tag{2.68}$$

This is a non-linear equation for $\boldsymbol{\Psi}$ which can be solved perturbatively. Let us expand J and $\hat{J}^{-1}$ for small $\boldsymbol{\Psi}$:

$$(\hat{J}^{-1})_{ij} = \delta_{ij} - \partial_j\Psi_i + \mathcal{O}(\Psi^2),$$
$$J = 1 + \sum_i \partial_i\Psi_i + \frac{1}{2}\sum_{ij}[\partial_i\Psi_i\,\partial_j\Psi_j - \partial_i\Psi_j\,\partial_j\Psi_i] + \mathcal{O}(\Psi^3).$$

The Zel'dovich approximation

To linear order, equation (2.68) reads

$$(\nabla\cdot\boldsymbol{\Psi})'' + \mathcal{H}(\nabla\cdot\boldsymbol{\Psi})' - \frac{3}{2}\mathcal{H}^2\Omega_{\rm M}(\nabla\cdot\boldsymbol{\Psi}) = 0. \tag{2.69}$$

This is exactly the equation for the linear density contrast, which we have already solved: $\nabla\cdot\boldsymbol{\Psi}(t,\mathbf{q}) \propto D_1^+(a)$. On the other hand, to linear order

$$\delta = \frac{1-J}{J} \simeq -\nabla\cdot\boldsymbol{\Psi}, \tag{2.70}$$

so the solution to (2.69) is

$$\nabla\cdot\boldsymbol{\Psi} = -\delta_0^{(1)}(\mathbf{q})D_1^+(a), \tag{2.71}$$

where we have only considered the growing mode.

We know that vortical modes ($\nabla\cdot\mathbf{v}=0$) decay rapidly ($\propto 1/a$), sow we may consider an irrotational $\boldsymbol{\Psi}$, in which case, in Fourier space:

$$\boldsymbol{\Psi}_{\mathbf{k}}(t) = -i\frac{\mathbf{k}}{k^2}\delta_{\mathbf{k}}(a). \tag{2.72}$$

Linear Lagrangian perturbation theory is usually referred to as the Zel'dovich approximation [97]. This model for structure formation allows an easy interpretation of the spatial distribution of matter in the Universe. Four different types of collapse can be defined in terms of the eigenvalues of the Jacobian $\partial_i\Psi_j$. Gravitational collapse creates spherical structures

where all the eigenvalues are equal and positive, filamentary objects arise where one of the eigenvalues is smaller than the rest, "pancakes" or "sheets" are formed where one eigenvalue is larger than the others, and voids appear where the three eigenvalues are negative. These are the Four elements of the so-called *cosmic web* classification of the matter density field [98].

The Zel'dovich approximation can be used to set up the initial displacements of the particles in N-body simulations, and currently there exist initial condition generators that use higher-order LPT for this task [99].

2.3.3 The halo model

There is a significant amount of empirical evidence that suggests that galaxies form preferably at high density regions, and that they reside inside large haloes of dark matter (e.g. [100]). Dark matter haloes are virialized clumps of dust that have departed the Hubble expansion through gravitational collapse. These structures have a highly non-linear nature, and cannot be studied through a perturbative analysis. However, the central role they play in the modelling of galaxy formation makes it necessary to find alternative approaches to learn about them. This study allows us also to understand many other problems in LSS: the nature of gravitational collapse in the very non-linear regime, the effect of environmental properties on the formation of structures, the mechanism that biases the abundance of tracers of the density field, etc.

Furthermore, the study of the properties and histories of dark matter haloes has motivated the development of a phenomenological model for the clustering of matter: the halo model [101, 102, 103]. The halo model has been proven to give accurate predictions when compared with N-body simulations, however its phenomenological nature requires that some of its ingredients must be tuned to these simulations. Nevertheless, the halo model is an extremely useful tool and we will describe its main assumptions and predictions here.

The halo mass function

The modelling of the halo mass function (the number density of dark matter haloes of different masses) is probably the most important ingredient of the halo model. As we will see, its derivation can be extended to understand other problems, such as the merging history of haloes or the halo bias. The first attempt at a theoretical estimation was made by Press & Schechter (PS) in their 1973 paper [104]. This calculation is based on the results of the spherical collapse model, linear perturbation theory, and the assumption that the linear density field is Gaussianly distributed.

We have already discussed linear perturbation theory, and details about the mathematics of Gaussian random fields can be found in appendix A. We have not described the spherical collapse model, but we will only need a couple of general results (a full description can be found in many text books, e.g. [19, 105]). This model studies the gravitational collapse of a spherical clump of dust. The high symmetry of this system makes it possible to follow the collapse analytically fully into the non-linear regime. According to this model, the size of the overdense region initially grows with the scale factor, but it eventually departs the Hubble expansion due to the gravitational attraction and collapses, reaching a singularity. This singularity would not occur in a practical case, and small departures from perfect sphericity and dissipating processes will create a virialized spherical structure. At this point the spherical collapse model predicts a central overdensity of $\Delta_{vir} \sim 178$, but more interestingly we can calculate the value that linear perturbation theory would predict at the time of collapse. This

is the δ_c parameter, which will be ubiquitous in what follows

$$\delta_c = \frac{3}{5}\left(\frac{3\pi}{2}\right)^{2/3} \simeq 1.686. \tag{2.73}$$

This value has been shown to depend very mildly on the background cosmological parameters [106], so it is safe to use 1.686 for all practical cases.

Now, consider the linear overdensity field smoothed over a comoving scale R with a top-hat filter (see section A.1.2). The PS model considers that a virialized structure forms whenever the linear overdensity field reaches the collapse threshold δ_c, and thus, assuming that this field is Gaussianly distributed, the number of points at which this occurs can be easily calculated. The mass of the halo corresponds to the mass of the region of space over which the field was smoothed

$$M = \frac{4\pi}{3}R^3\bar{\rho} = 1.16\times 10^{12}\Omega_{\rm M}\left(\frac{R}{h^{-1}\,{\rm Mpc}}\right)^3 (M_\odot\, h^{-1}). \tag{2.74}$$

For a Gaussian field, the probability of exceeding δ_c is

$$P(\delta_R \geq \delta_c) = \frac{1}{2}\left[1 - {\rm erf}\left(\frac{\delta_c}{\sqrt{2}\sigma_R}\right)\right],$$

where σ_R^2 is the variance of the smoothed field and

$${\rm erf}(x) \equiv \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt.$$

Thus, according to the PS model, the fraction of the volume of the Universe that has collapsed into structures of mass larger than M (i.e. sizes larger than $R(M)$) is

$$F(>M) = \frac{1}{2}\left[1 - {\rm erf}\left(\frac{\delta_c}{\sqrt{2}\sigma_{R(M)}}\right)\right], \tag{2.75}$$

Taking $R\to 0$ (i.e.: $\sigma\to\infty$) we find that only half of the mass of the Universe has collapsed into virialized structures, which is half of what we would expect. The reason for this is that we have only taken into account regions that were initially overdense, and not underdense regions embedded in larger overdensities. In the original paper by Press & Schechter they argue about this factor 2 handwavingly and incorporate it by hand, but we will see a proper derivation of this factor later on. Thus we obtain the PS formula

$$F(>M) = 1 - {\rm erf}\left(\frac{\nu_c}{\sqrt{2}}\right), \quad \nu_c \equiv \frac{\delta_c}{\sigma_{R(M)}} \tag{2.76}$$

From the cumulative collapsed fraction $F(>M)$ one can immediately calculate the number density of collapsed objects of a given mass. Let us define three useful quantities:

- **Collapsed fraction, $F(>M)$:** fraction of the Universe that has collapsed into haloes of mass larger than M.
- **Multiplicity function $f(M)$:** fraction of the Universe that has collapsed into haloes in a unit logarithmic mass interval

$$f(M) = -\frac{dF(>M)}{d\log M} \equiv g(\sigma)\left|\frac{d\log\,\sigma(M)}{d\log\,M}\right|. \tag{2.77}$$

- **Mass function, $n(M)$:** number density of haloes of mass $M \in (M, M + dM)$.

$$n(M)\, dM \equiv -\frac{\rho_{\rm M}}{M}\frac{dF(>M)}{d\log M} d\log M = \frac{\rho_{\rm M}}{M^2}\, g(\sigma) \left|\frac{d\log\, \sigma(M)}{d\log\, M}\right| dM. \tag{2.78}$$

Press & Schecther found that, when written in terms of σ (or ν_c), the function g defined above is **universal** (i.e. cosmology-independent) with the form:

$$g_{PS}(\nu_c) \equiv \sqrt{\frac{2}{\pi}}\nu_c\, \exp\left(-\frac{\nu_c^2}{2}\right).$$

The requisite of completeness (i.e. $F(>M \to 0) \to 1$) is granted provided that

$$\int_0^\infty g(\nu)\frac{d\nu}{\nu} = 1. \tag{2.79}$$

As must be evident, the PS theory is not mathematically rigorous: first of all there is the factor 2 fuss. Also, we know that the overdensity can only be regarded as Gaussian in the linear regime (for instance, it cannot be $\delta < -1$). Furthermore, we are trying to obtain the number of collapsed (non-linear) objects from the linear density contrast. Nevertheless this treatment gives a good qualitative fit to the result obtained from N-body simulations, even though it fails to reproduce the details. A hint about the possible explanation for this failure, as well as a rigorous derivation of the factor 2 can be found within the excursion-set formalism, which we will describe shortly. In any case, a general result can be obtained from the PS model: the mass function should be a universal function of the variance σ_R. This is a sound result, since the statistics of a Gaussian field are fully determined by its variance. Several groups have worked out alternative parametrizations of the mass function (some based on extensions of the PS model), and this quantity can now be predicted for a large range of cosmological models with a precise fit to the simulated data up to fairly large masses [107, 108, 109]. A popular parametrization was given by Sheth & Tormen [110] based on the ellipsoidal collapse model:

$$\begin{aligned} g_{ST}(\nu_c) \equiv\ & A\sqrt{\frac{2\,a}{\pi}}\left[1 + \left(\frac{1}{a\,\nu_c^2}\right)^p\right]\, \nu_c\, \exp\left(-a\,\frac{\nu_c^2}{2}\right), \\ (A, a, p) =\ & (0.322, 0.707, 0.3). \end{aligned} \tag{2.80}$$

The excursion set: random walks, the factor 2

It is possible to make sense of the missing factor 2 in the PS result through the so-called *excursion set* formalism [111]. It is a very useful exercise to go through the details of this derivation, since it provides a way to calculate conditional mass functions quite straightforwardly.

Let $\delta(\mathbf{x}, R)$ be the density contrast filtered on scales below R at a given fixed point $\mathbf{x}$. By following the value of δ at this point for different filters, we will define a trajectory in the (R, δ) plane. In order to simplify the analysis, we will use Heavyside window functions *in Fourier space* (i.e. $W(kR) = \Theta(kR)$, also called a sharp-k filters). Asymptotic homogeneity implies that $\delta(\mathbf{x}, R \to \infty) \to 0$, therefore all these trajectories start from 0 at $R = \infty$. Furthermore, going from a bigger filtering scale R_1 to a smaller R_2 with our choice of $W(kR)$ implies adding to δ a set of uncorrelated Fourier modes. Since these have Gaussian statistics, the probabilities that this contribution will be positive or negative are the same. We are interested in the number of trajectories that have crossed the threshold δ_c for the first time at some filtering scale $R_{\rm cross}$ larger than R. On the one hand, all the trajectories that end

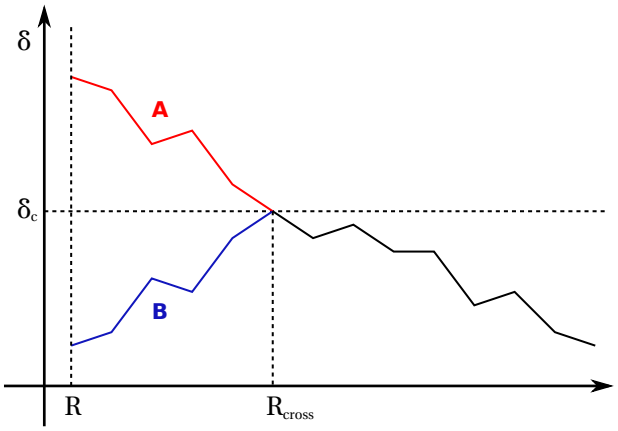

Figure 2.3: Illustration of the excursion-set formalism. Random walks in the $\delta - R$ plane generated using sharp-k filters are equally likely to end up above (A) or below (B) δ_c once the cross the threshold.

up above δ_c satisfy this requirement (since they all started from $\delta = 0$). For each of these trajectories (A), we can construct another one (B) by mirroring with respect to the threshold the part of the A corresponding to scales smaller than $R_{\rm cross}$ (see fig. 2.3), and these mirror walks will also satisfy our requirement. For our choice of window function these two sets of walks are equally probable and, furthermore, all the trajectories we are looking for can be classified into one of these two types. Therefore, the (normalized) number of trajectories that have crossed the threshold for filtering scales above R is just twice the number of those that end up above the threshold, and hence the missing factor 2:

$$P(\delta(\mathbf{x}, R') = \delta_c, R' > R) = 2\, P(\delta(\mathbf{x}, R) > \delta_c) = 1 - \mathrm{erf}\left(\frac{\nu_c}{\sqrt{2}}\right)$$

Notice that this result depends dramatically on the type of window function we have chosen (sharp-k). For other filters, the adjacent steps of each walk will not be uncorrelated, and steps going up or down will not be equally probable. In fact, the excursion set formalism can be recreated using walks with correlated steps [112, 113], and in this case one obtains the PS result *without* the factor 2 for high masses and a surplus of collapsed objects at low masses.

Merger trees

We know that haloes collapse in a "bottom-up" picture, with lower-scale structures forming first and then collapsing into haloes of larger size. To study this hierarchical collapse it would be useful to find the fraction of the mass collapsed into structures of mass M_1 at z_1 that was in structures of mass $> M_2$ (with $M_2 < M_1$) at $z_2 > z_1$. This problem was studied in [114, 115] and can be translated into the excursion set formalism: we are looking for the number of trajectories that, starting from $\delta_1 = \delta_c/D(z_1)$ at a filtering scale $R_1 = R(M_1)$, end up above $\delta_2 = \delta_c/D(z_2)$ at $R_2 = R(M_2)$ (note that they must start from δ_1 and not above, since we are interested in final structures of mass $M = M_1$ and not $M > M_1$). Here $D(z)$ is normalized to $D(z = 0) = 1$ (we are thinking of $\delta(\mathbf{x}, R)$ as a static quantity and we have translated all the time dependence to the threshold).

The conditional mass function. In general, we want to calculate

$$P(\delta(\mathbf{x}, R') = \delta_2, R' > R_2 | \delta(\mathbf{x}, R_1) = \delta_1).$$

This problem is commonly known as the *conditional mass function*, and can be tackled within the excursion-set framework. As a first step let us find the probability that a trajectory starting at δ_1 at R_1 will end up at $\delta \pm d\delta$ on $R < R_1$:

$$P(\delta(\mathbf{x}, R) = \delta | \delta(\mathbf{x}, R_1) = \delta_1)\, d\delta = \frac{P(\delta(\mathbf{x}, R) = \delta, \delta(\mathbf{x}, R_1) = \delta_1)\, d\delta\, d\delta_1}{P(\delta(\mathbf{x}, R_1) = \delta_1)\, d\delta_1}.$$

For a Gaussian field this is

$$P(\delta(\mathbf{x}, R) = \delta | \delta(\mathbf{x}, R_1) = \delta_1)\, d\delta = \frac{1}{\sqrt{2\pi(1-\varepsilon^2)}} \exp\left(-\frac{1}{2}\frac{(\nu - \varepsilon\nu_1)^2}{1-\varepsilon^2}\right) d\nu,$$

where

$$\nu \equiv \frac{\delta}{\sigma(R)}, \quad \nu_1 \equiv \frac{\delta_1}{\sigma(R_1)}, \quad \varepsilon \equiv \frac{\langle \delta_R \delta_{R_1} \rangle}{\sigma(R)\sigma(R_1)},$$
$$\langle \delta_R \delta_{R_1} \rangle = \frac{1}{2\pi^2} \int_0^\infty W(kR) W(kR_1)\, P_k\, k^2\, dk.$$

Note that the window functions for R and R_1 need not be of the same type, although we have assumed so here.

For sharp-k filters and since $R < R_1$ we have $W(kR)\, W(kR_1) = W(kR_1)^2$, therefore $\langle \delta_R \delta_{R_1} \rangle = \sigma^2_{R_1} \Rightarrow \varepsilon = \sigma(R_1)/\sigma(R)$. Following the reasoning used before for the unconditional mass function, our conditional mass function will be given by twice the normalized number of these trajectories that end up above the second threshold δ_2 at R_2. Thus

$$F(> M_2, z_2 | M_1, z_1) = 1 - \mathrm{erf}\left(\frac{\nu_{12}}{\sqrt{2}}\right), \tag{2.81}$$

where

$$\nu_{12} = \frac{\delta_2 - \delta_1}{\sqrt{\sigma^2_{R_2} - \sigma^2_{R_1}}} = \frac{\delta_c}{\sigma_{R_2} D(z_2)} \frac{1-t}{\sqrt{1-\varepsilon^2}}, \quad t \equiv \frac{D(z_2)}{D(z_1)}.$$

Eq. (2.81) gives the fraction of the volume collapsed into haloes of mass M_1 at redshift z_1 that has collapsed into haloes of mass $> M_2$ at z_2. From this it is easy to calculate the mean number of progenitors with mass $M_2 \pm dM_2/2$ at $z > z_1$ of a halo of mass M_1 at z_1:

$$N(M_2, z | M_1, z_1) dM_2 = \frac{M_1}{M_2} \left| \frac{dF(> M_2, z | M_1, z_1)}{dM_2} \right| dM_2. \tag{2.82}$$

In terms of the universal function $g(\nu)$ this is

$$N(M_2, z | M_1, z_1) dM_2 = \frac{M_1}{M_2} g(\nu_{12}) \frac{1}{1-\varepsilon^2} \left| \frac{d \log \sigma_2}{d \log M_2} \right| \frac{dM_2}{M_2}. \tag{2.83}$$

Notice that we get the same function we got in the unconditional case with the old ν variable replaced by ν_{12}. As we saw, the PS prediction for the *unconditional* mass function was not very good in the details and better parametric fits have been developed. Although it is a tempting alternative, it has been shown that using the rescaling $\nu \to \nu_{12}$ in these parametrizations does not give a good approximation to the *conditional* mass function when studying halo mergers [116].

Spatial clustering and bias

Since the PS model predicts that haloes form whenever the density contrast exceeds the threshold δ_c, it is reasonable to think that haloes will be relatively more abundant in regions of the Universe with higher large-scale densities. This phenomenon is called *halo biasing*, and can also be quantified using the conditional mass function. As we just saw The probability for (δ_1, R_1) conditional to (δ_2, R_2) is:

$$P(\delta_1, R_1|\delta_2, R_2)d\delta_1 = \frac{d\nu_1}{\sqrt{2\pi(1-\varepsilon^2)}} \exp\left(-\frac{1}{2}\frac{(\nu_1 - \varepsilon\nu_2)^2}{1-\varepsilon^2}\right), \tag{2.84}$$

where $\nu_i \equiv \delta_i/\sigma(R_i)$ and ε was defined above. Here we are thinking of R_2 as the scale with which we describe the local environmental density and R_1 as the scale related to the mass of the haloes residing in those environments. Thus for our purposes $R_2 > R_1$ and $|\delta_2| < \delta_c$. Following the same steps as before, the fraction of the volume in regions where $\delta_{R_{\rm env}} = \delta_{\rm env}$ that has collapsed into haloes of masses $> M$ is, in the excursion set formalism

$$F(>M|\delta_{\rm env}, R_{\rm env}) = 1 - {\rm erf}\left[\frac{1}{\sqrt{2}}\frac{\nu_c - \varepsilon\nu_{\rm env}}{\sqrt{1-\varepsilon^2}}\right] \tag{2.85}$$

From this result it is easy to calculate the environmental mass function $n(M, |\delta_{\rm env}, R_{\rm env})\, dM$.

Linear bias. In the approximation $R_{\rm env} \gg R(M)$ and $|\delta_{\rm env}| \ll \delta_c$, and assuming sharp-$k$ filters we have

$$\frac{\nu_c - \varepsilon\nu_{\rm env}}{\sqrt{1-\varepsilon^2}} \simeq \frac{\delta_c - \delta_{\rm env}}{\sigma(R(M))}. \tag{2.86}$$

Hence it is easier to reach the collapse threshold in overdense regions, since the effective threshold is reduced to $\delta_c - \delta_{\rm env}$, and, as we had predicted, more haloes will be formed. This result is the so-called *peak-background split*, and is the basis of our current understanding of halo and galaxy biasing. These results were first explored by Mo & White [76] and have since been refined using different approaches (e.g. [117]).

The Lagrangian halo number density contrast is

$$\delta_n^L(M, \delta_{\rm env}) = \frac{n(M|\delta_{\rm env}, R_{\rm env})}{n(M)} - 1. \simeq \frac{\nu_c^2 - 1}{\delta_c}\,\delta_{\rm env} \tag{2.87}$$

(where we have linearized in $\delta_{\rm env}$). In Eulerian space this is

$$\delta_n(M, \delta_{\rm env}, z) = b(M, z)\,\delta_{\rm env}(z) \tag{2.88}$$

with $\delta_{\rm env}(z) = \delta_{\rm env}(0)D(z)$ and

$$b(M, z) = \frac{\nu_c^2 - 1}{\delta_c\, D(z)} + 1. \tag{2.89}$$

$b(M, z)$ is the so-called *halo bias*, and represents the enhancement (or inhibition) of halo formation in regions with different background densities. Defining M_* as $\sigma(R(M_*)) = \delta_c$ we see that haloes with $M < M_*$ are positively biased, while those with $M > M_*$ are anti-biased.

Assuming that the linear bias relation (2.88) holds locally (see **caveats** below), the two-point correlation function of the halo number density is related to the matter correlation function $\xi(r, z)$ via

$$\begin{aligned}\xi_h(r, z|M_1, M_2) &\equiv \langle\delta_n(\mathbf{x}, z, M_1)\delta_n(\mathbf{x}+\mathbf{r}, z, M_2)\rangle \\ &= b(M_1, z)\, b(M_2, z)\, \xi(r, z)\end{aligned} \tag{2.90}$$

Finally, just like the PS mass function, this bias function is not precise in the details. This can be improved using the ellipsoidal collapse (Sheth & Tormen's mass function, eq. (2.80)), which yields

$$b(M,z) = 1 + \frac{1}{\delta_c\, D(z)}\left[\nu_c^2 + b\nu_c^{2(1-c)} - \frac{1}{\sqrt{a}}\frac{\nu_c^{2c}}{\nu_c^{2c} + b\,(1-c)\,(1-c/2)}\right], \tag{2.91}$$

with $(a,b,c) = (0.707, 0.5, 0.6)$.

Caveats. The results we have just derived are based on very strong assumptions which are, in general, false. We will address two of them here:

- **Nonlinear bias:** the linear bias relation (2.88) is not true for all $\delta_{\rm env}$, and $R_{\rm env}$. Deviations from it may come from non-negligible $\delta^2_{\rm env}$ terms or from mass scales $R(M)$ comparable to the scale of the environment $R_{\rm env}$, and in general we have

$$\delta_n = \delta(M,z,\delta_{\rm env}) = \sum_{k=0}^{\infty}\frac{b_k(M,z)}{k!}\,\delta^k_{\rm env}. \tag{2.92}$$

  For a general dependence like this, one can always relate the moments of the distribution of δ_n to those of $\delta_{\rm env}$ [118]. Nonlinear bias effects will be most evident when the assumption $|\delta_{\rm env}| \ll 1$ breaks down, and therefore the linear approach for the correlation function will be a good approximation for large scales, where $\xi(r) \ll 1$ (note however that this is not so in the presence of primordial non-Gaussianities [119]).

- **Stochastic bias:** the derived linear bias relation, and even the general non-linear one above, need not hold locally, but as an average relation. In general the halo number density contrast will not be deterministically related to the local environmental matter density. This stochasticity can be modelled as [120]

$$\delta_n(\mathbf{x},z,M) = b(M,z)\,\delta(\mathbf{x},z) + \epsilon(\mathbf{x}), \tag{2.93}$$

  where $\epsilon(\mathbf{x})$ is a zero-mean stochastic field ($\langle\epsilon(\mathbf{x})\rangle = 0$). With this, the two-point correlation function is

$$\begin{aligned}\xi_h(r,z|M_1,M_2) &\equiv \langle\delta_n(\mathbf{x},z,M_1)\,\delta_n(\mathbf{x}+\mathbf{r},z,M_2)\rangle \\ &= b(M_1,z)\,b(M_2,z)\,\xi(r,z) + \\ &\quad + (b(M_1,z)+b(M_2,z))\,\langle\epsilon(\mathbf{x})\,\delta(\mathbf{x}+\mathbf{r})\rangle + \\ &\quad + \langle\epsilon(\mathbf{x}+\mathbf{r})\,\epsilon(\mathbf{x})\rangle.\end{aligned} \tag{2.94}$$

  Thus only if the stochastic term is uncorrelated at long distances will the linear bias relation for the correlation function hold on large scales.

Internal structure of dark matter haloes

Modelling haloes as spherical objects, the matter distribution can be described by a density profile $\rho(r)$. There exist different models of spherical collapse that predict different density profiles. However, fully modelling the physics involved is extremely challenging, and a fit to numerical N-body simulations is usually the best option. Two main types of profiles have been advocated in the literature

- **NFW.** The Navarro-Frenk-White (NFW) profile [121] depends on two parameters that govern the amplitude and shape of the profile:

$$\rho_{\rm NFW}(r) = \rho_{\rm crit}\frac{\delta_{\rm char}}{r/r_s(1+r/r_s)^2}. \tag{2.95}$$

- **Einasto.** The Einasto profile [122] has three parameters, governing the amplitude, shape and tilt:

$$\rho_{\rm Ein}(r) = \rho_{-2}\exp\left(-\frac{2}{\alpha}\left[\left(\frac{r}{r_{-2}}\right)^{\alpha}-1\right]\right). \tag{2.96}$$

More recent simulations (than the ones used by NFW) have shown systematic deviations from NFW and that Einasto may give a better fit (even fixing $\alpha = 0.17$) [123], however the matter is still not settled. There is also the problem of the central cusp in the NFW profile ($\rho_{\rm NFW}(r \to 0) \propto 1/r \to \infty$), however it must be noted that although the Einasto profile is mathematically not "cuspy", numerically it is. In any case, individual haloes usually differ more (depending on their merging history) from either NFW or Einasto than the two profiles do from each other.

Wrapping up: The halo model

In the halo model one thinks of the dark matter content of the Universe as a homogeneous background density on top of which sit haloes with a universal density profile $\rho_h(\mathbf{x}) = M\,u(\mathbf{x}|M)$, extending over large scales from halo to halo. Here u is normalized to unity

$$\int dx^3 u(\mathbf{x}|M) = 1. \tag{2.97}$$

Hence, knowing the abundance of haloes of different mass, their density profiles and the way in which their number density is biased with respect to the large-scale matter density, it is possible to model the power-spectrum on non-linear scales in terms of the linear power spectrum and the halo properties. This is the main result of the halo model, and its derivation is detailed in several references (e.g. [105]). We will only quote here its final form.

The power spectrum has two main contributions, coming from pairs of matter elements belonging to different haloes (the 2-halo term) and to the same halo (the 1-halo term).

$$P(k) = P_{\rm 1h}(k) + P_{\rm 2h}(k), \tag{2.98}$$

with

$$P_{\rm 1h}(k) = \frac{1}{\bar\rho^2}\int dM\, M^2 n(M)\,|u(k|M)|^2, \tag{2.99}$$

$$P_{\rm 2h}(k) = P_{\rm lin}(k)\left[\frac{1}{\bar\rho}\int dM\, M\, n(M)\, b(M)\, u(k|M)\right]^2. \tag{2.100}$$

The halo model as has been presented can reproduce the main features of the non-linear power spectrum and is a very useful tool to understand the non-linear scales. It has also recently been used to calculate the effect of primordial non-Gaussianities in the bispectrum [124]. However, these predictions are not accurate in the details, and a fitting formula for the non-linear power spectrum has been developed [125] based on the main results of the halo model. This fitting was implemented in a code called `HALOFIT`, which is widely used within the cosmology community.

2.4 Non-linearities II: numerical simulations

Although the analytical and phenomenological methods outlined in the previous section are extremely useful, only approximate results can be extracted from them, and in order to study the exact evolution of the density field we must resort to the *brute force* solution of N-body simulations, which are the topic of this section. In an N-body simulation the density field is represented by a set of fictitious discrete particles interacting among themselves. For LSS purposes these particles usually represent the dark matter density field and therefore only interact gravitationally. These particles are given initial comoving coordinates $\mathbf{q}$ and the background expansion is taken into account by scaling them at every time iteration with the scale factor to get the physical coordinates $\mathbf{x} = a\,\mathbf{q}$.

It is important to note that these simulations are based on Newtonian physics. The equations of motion for each particle are

$$\frac{d\mathbf{q}}{dt} = \mathbf{u}, \quad \frac{d\mathbf{u}}{dt} = -2H\,\mathbf{u} - \frac{1}{a^2}\nabla\Phi, \tag{2.101}$$

where $\mathbf{u}$ is the comoving velocity, H is the expansion rate and Φ the Newtonian potential on the particle. The mechanism is, in principle, very simple: at each time step the comoving position of a particle is varied by $d\mathbf{q} = \mathbf{u}\,dt$ and its velocity by $d\mathbf{u} = \dot{\mathbf{u}}\,dt$, with $\dot{\mathbf{u}}$ given by the equations above. However there are a few complications:

- How do we set the initial positions and velocities for the particles, so that they represent a density field with physical inhomogeneities?

- If there are N particles, in principle one has to calculate the potential on each of them created by the other $N-1$ particles. Thus the computational time for this algorithm scales as N^2 with the number of particles. There are several methods which can be used to deal with this problem that we will describe below.

Besides these problems, N-body simulations are not all-powerful, and have other inherent shortcomings:

- They are Newtonian and in principle some general-relativistic effects (e.g. horizon effects) could be overseen.

- The particles evolve in a box of large but finite size (tens or hundreds of Mpc), usually with periodic boundary conditions, and we cannot explore scales beyond the box size $L_{\rm box}$. For sufficiently large sizes this is not a big problem, since large scales remain in the linear regime until the present day.

- Since the total mass of the Universe must be contained in the N simulation particles, together with $L_{\rm box}$, this determines the mass of our particles:

$$m_p = \bar{\rho}\,\frac{L_{\rm box}^3}{N}. \tag{2.102}$$

  This implies that we will only be able to resolve structures of masses larger than a mass threshold of a few m_p's.

- These simulations are computationally costly and with each one of them we can study the evolution of the density field in only one particular cosmological background.

2.4.1 Initial conditions

N-body simulations are initialized at an early cosmic time t_i (usually $z_i \sim \mathcal{O}(10-100)$). By placing $N_{\rm side}^3$ particles at rest at the nodes of a comoving grid with lattice spacing $a = L_{\rm box}/N_{\rm side}$ at this starting redshift and allowing them to evolve, we would simulate a universe with a perfectly homogeneous density $\rho = m_p(N_{\rm side}/L_{\rm box})^3$, since the total gravitational force on each particle would cancel exactly. In order for these particles to form inhomogeneities, we must displace them from these nodes in the initial conditions. Hence we need to answer the question: how far away from their "primordial" comoving positions are the particles of the cosmic fluid at this starting time? This question can be easily answered using Lagrangian perturbation theory, which we described in the previous section.

At linear order in LPT, the Zel'dovich approximation relates the initial comoving coordinates $\mathbf{q}_i$ of a particle in a perturbed homogeneous universe with its final one at a later time $\mathbf{q}(t)$ by assuming that once the particle starts to depart from its initial comoving position (the one it would always occupy in a perfectly homogeneous universe), it does so in the same direction all the time. This relation is

$$\mathbf{q}(z) = \mathbf{q}_i - i\,D(z)\sum_{\mathbf{k}}\frac{\delta_{\mathbf{k}}}{k^2}\,\mathbf{k}\,e^{i\mathbf{k}\cdot\mathbf{q}}. \tag{2.103}$$

Thus the procedure to set the initial condition is to generate two Gaussian random numbers $(a_{\mathbf{k}}, b_{\mathbf{k}})$ for each $\mathbf{k}$ with mean 0 and dispersion $\sigma^2 = \frac{1}{2}\left(\frac{L_{\rm box}}{2\pi}\right)^3 P(k)$ so that $\delta_{\mathbf{k}} = a_{\mathbf{k}} + ib_{\mathbf{k}}$ and then apply eq. (2.103) with $z = z_{\rm start}$ and $\mathbf{q}_i$ the unperturbed positions. It has been shown that without second and higher-order terms in LPT, the Zel'dovich approximation can excite long-lived transients in the evolution of the density and velocity fields. For this reason, algorithms using second-order Lagrangian perturbation theory have been developed [99], which need a much smaller starting redshift to attain the same precision.

2.4.2 Codes

Once the initial conditions have been set, the gravitational evolution is calculated by integrating the equations of motions for each of the particles. As we observed at the beginning of this section this is a non-trivial computational challenge, and several algorithms have been developed to deal with the N^2 scaling problem. Here are the most popular ones:

1. **PP.** The particle-particle (PP) method is the naive one. The force on each particle is calculated by summing the gravitational force caused by each of the other $N-1$.

2. **PM.** Particle-mesh (PM) codes use a mesh and calculate, in every mesh cell, the density. Then the potential is calculated solving Poisson's equation using a Fast Fourier Transform. As a result their resolution is limited by the lattice spacing. They are, however, the fastest method and are able to deal with larger numbers of particles.

3. **P^{3}M.** Particle-particle, particle-mesh (P^{3}M) combine the two previous methods. The PM part takes care of large-scale forces and the PP part computes the contribution from particles at small distances. When there is strong clustering this becomes less efficient, since the PP part dominates. A good solution to this is to use refined grids in high-density areas.

4. **Tree.** Tree codes arrange the particles in different hierarchical groups. When the force on a given particle is calculated, the contribution from distant groups is approximated

by their lowest multipole moments. This reduces the scaling of the computational time to $\mathcal{O}(N \log N)$. Usually particles are grouped by recursively dividing the box into 8 sibling cubes (oct-tree) of half the size. The size of the group of siblings with which to compute the force on a given particle must satisfy the following opening criterion:

$$l < \theta\, r, \tag{2.104}$$

where r is the distance from the particle to the centre of mass of the group and θ is the opening parameter.

5. **TreePM.** TreePM codes combine the Tree and PM algorithms. The PM part is used in Fourier space to calculate the gravitational force on very large scales, while the Tree is used for short-range forces. One of the most widely used codes for cosmological N-body simulations, `GADGET-2` [126] uses a TreePM algorithm, although the PM part may be switched off if so desired.

6. **AMR.** The main caveat of pure PM methods, their limited spatial resolution, can be overcome with the Adaptive Mesh Refinement method, in which the mesh resolution is improved in high density regions. These methods are also very useful to compute the hydrodynamical forces in an Eulerian framework (its Lagrangian counterpart being the Smoothed Particle Hydrodynamics method - SPH).

In all these cases a softening of the Newtonian $\propto 1/r$ gravitational potential is needed to suppress large-angle scattering in two-body collisions. This effectively introduces a cut-off given by the softening length h. For a given h it is important to choose a large enough particle number so that relaxation effects are suppressed, otherwise the N-body system would not reproduce a collisionless system reliably. The choice of the optimal softening length is still an open issue [127], but a common choice is $h \sim (30 - 50)\, L_{\rm box}/N_{\rm side}$.

The topic of cosmological simulations is extremely involved and we have not attempted a thorough description here. There exist multiple reviews and textbooks in the literature [128, 129, 130] that the reader is referred to for further details.

CHAPTER 3

Model-independent determination of the BAO scale in redshift surveys

As has already been described in previous chapters, the sound waves in the baryon-photon plasma at early times, caused by pressure gradients, leave their imprint in the late-time spectrum of matter perturbations. This signature takes the form of an oscillatory contribution to the power spectrum in Fourier space, and a peak or bump in the 2PCF in configuration space. The frequency of these oscillations and the position of the peak correspond to the size of the sound horizon at the baryon drag epoch r_s: the maximum distance that a sound wave may have travelled before baryons decoupled and pressure gradients ceased to be relevant. The sound horizon scale can then be used as a standard ruler, and by measuring the BAO at later times one can give constraints on the distance-redshift relation and the expansion history of the Universe.

The BAO feature has been measured at late times in galaxy redshift surveys, providing an excellent addition to the high redshift measurement from the CMB. The standard way of performing this measurement is detailed in section 3.1. The caveats and possible shortcomings of the standard method, described below, have motivated our work in this subject. We have proposed an alternative approach, which enables us to measure the BAO in the angular and radial directions separately using only pure observables. This method is described in sections 3.2 and 3.3, and we compare it with the conventional approach in section 3.4. There exist excellent reviews in the literature which provide an in-depth description of the BAO as an observational probe [131, 132].

3.1 Standard approach to baryon acoustic oscillations

The baryon acoustic oscillations scale was first measured from a galaxy redshift survey in 2005 by the SDSS collaboration [133] in their 3rd Data Release (see also [134] for a similar detection by the 2dFGRS collaboration in their final dataset). Since then, these measurements have been improved and new detections have been claimed (e.g. [25, 26, 27, 28, 30]). In all cases the analysis has been performed in a similar way, which we will describe here.

3.1.1 Galaxy redshift surveys

Although an in-depth description of the technicalities of galaxy redshift surveys, their challenges and future prospects, is well beyond the scope of this work, we would like to briefly point out some of the key details of these surveys which will be important for the discussion in the rest of this section.

In a galaxy redshift survey three observable quantities locate a galaxy in space: its angular coordinates, for instance right ascension (RA, α) and declination (dec, δ) in equatorial coordinates, and its redshift z. The redshift is measured through the shift in the emission spectra of the observed galaxies caused by the expansion of the Universe, which dilates the wavelength of the emitted photons. It is crucial to note that, although the redshift contains information about the distance to the observed object, it does not provide an actual distance measurement without further assumptions. In order to translate this redshift into a distance, a particular cosmological model must be adopted, which provides a specific redshift-distance relation.

For current and future surveys, there exist two different approaches to measuring z:

- **Spectroscopic redshifts.** In spectroscopic surveys the full spectrum of each galaxy is measured and used to determine its redshift with a very good precision. The broadband photometry of all observed galaxies is first determined and the different populations are characterized. The target galaxies are then selected and located in the sky. A fibre is dedicated for each selected source and measures its spectrum by separating the light in many narrow bins of wavelength. Since each bin measures only a very small portion of the light, long integration times are needed for a good signal-to-noise ratio.

  This process is very time-consuming, and can only be done for a subset of the observed sample. At low redshifts, where the available volume is small and statistical uncertainties are dominated by cosmic variance, this may not be an issue. However, at higher redshifts ($z \geq 1$) the cosmic variance is suppressed for the same sky fraction, and shot noise becomes a more significant concern.

  The most important BAO measurements to date have been performed on spectroscopic surveys, such as 2dFGRS [135], 6dF [136], WiggleZ [137] or SDSS [138].

- **Photometric redshifts.** The redshift may also be determined from the galaxy's photometry: the magnitudes in a few broadband filters. There exist different techniques for doing this, like the fitting of redshifted spectral templates (see [139] and [140] for a Bayesian implementation) or the use of a training set of spectra in a neural network [141]. The resulting photometric redshift (photo-z hereon) has a larger error σ_z than a spectroscopic redshift. This error can usually be parametrized as a function of redshift as $\sigma_z = \sigma_0 (1 + z)$ due to the stretching of the spectra with redshift for a filter set of constant resolution. The magnitude of this error depends on the filter set, the type of galaxy and the method used, and, for most of the surveys listed below, is not expected to be below $\sigma_0 \sim 0.03$. Due to this large error, structures are washed off along the line of sight, and much of the radial information is lost. The use of photo-z's also entails having to deal with new sources of systematic errors, and the analysis is more complex than in the spectroscopic case.

  In spite of these caveats, photometric redshifts are expected to be extremely useful for LSS in the future. Since full spectra are not measured, the redshift can be estimated for all the observed sample (with better or worse accuracy), and therefore the shot noise problem at high redshifts is largely alleviated [142].

The Dark Energy Survey (DES) [143] and PanSTARRS [144] are two near-term large area photometric redshift surveys which will be able to measure the BAO at $z \gtrsim 1$ with unprecedented accuracy, and in the future the LSST [145] should be able to get to $z \sim 2$. Two Spanish projects, PAU [146] and JPAS [147] will aim to obtain photometric redshifts using medium-band filters, an intermediate approach between spectroscopy and photometry, which will improve the photo-z accuracy significantly.

Besides redshifts and angular positions, measuring the photometry of the survey's galaxies is a crucial task. For example the colours or the relative magnitudes in a given band, contain information about the galaxy type, and this information may be used to detect sources of systematic uncertainties and to minimize their effect, as occurs for example with the galaxy clustering bias. Other galaxy properties may be used as observational probes for cosmology (e.g. galaxy shapes and sizes for Weak Lensing). For the purpose of the forthcoming discussion, we will mainly be interested in the angular position and redshift determination. There are many other observational challenges that influence the constraining power of galaxy surveys: survey completeness, determining the redshift distribution, star-related effects, etc. However, discussing all of these would is not directly relevant for our work.

3.1.2 Measuring the BAO from the correlation function monopole

Baryon Acoustic Oscillations manifest themselves as a sinusoidal contribution in the power spectrum and as a bump in the configuration-space 2PCF. The origin of this signature has been described before in chapter 2, and the correspondence between Fourier and configuration spaces is easy to understand, given that $\xi(r)$ and $P(k)$ are Fourier transforms of one another, and the Fourier transform of a sinusoid with frequency ω is a Dirac's delta function at ω. The width of the BAO peak is due to the damping of the oscillations in the $P(k)$ caused by different effects.

Up until now, the most important BAO measurements have been drawn from the power-spectrum and the correlation function monopoles (angle-averaged quantities in redshift space). We will focus here on the measurement from the 2PCF, but the technique for the $P(k)$ measurement is very similar (e.g. see [30]). The monopole makes the most efficient use of the data in a survey in terms of shot noise, since it uses all available galaxy pairs. The drawback of using the angle-averaged 2PCF is that the radial and angular scales, which contain complementary information, are mixed. The method described here corresponds to the one used in the latest measurement of the acoustic peak by the BOSS collaboration [30], which is very similar to the methods used before.

The 2PCF is calculated from the galaxy distribution following the method described in appendix A.2.2. It is important to stress the fact that in order to estimate $\xi(r)$, which depends on the 3-dimensional distance r, the redshifts of the survey galaxies must be translated into comoving distances $\chi(z)$ using equation (1.17), for which a fiducial cosmological model must be adopted. In order to locate the position of the BAO peak, a fit is made on the measured 2PCF. The fitting model is based on a rescaled template of the theoretical correlation function monopole corresponding to the fiducial cosmology and the rescaling parameter contains the information about the peak position. The specific model used by [30] is

$$\xi_{\rm fit}(r) = B^2\, \xi_{\rm th}(\alpha\, r) + A(r), \tag{3.1}$$

where $\xi_{\rm th}$ is the theoretical prediction for the fiducial model, B^2 takes care of anything affecting the overall normalization, such as a large-scale bias factor, and α is the rescaling parameter. The function $A(r)$ absorbs any unmodelled broadband features of the correlation function,

coming for example from non-linearities on small scales, incorrect modelling of RSDs, scale-dependent bias or errors in the determination of the survey redshift distribution. The choice of $A(r)$ in [30] is

$$A(r) = a_0 + \frac{a_1}{r} + \frac{a_2}{r^2}. \tag{3.2}$$

It is important to model the theoretical correlation function as accurately as possible, so that the fit does not depend much on the nuisance parameters a_i. It is usually computed from the theoretical power spectrum using eq. (A.25) and effects such as non-linearities or RSDs are included in the $P(k)$. In particular it is customary to include the damping of the BAO wiggles due to non-linear structure formation (see section 3.1.4 below)

$$P_{\rm NL}(k) = P_{\rm L}(k)\, e^{-\sigma_v^2 k^2/2}, \quad \sigma_v^2 \equiv \frac{1}{6\pi^2}\int_0^\infty P_{\rm L}(k)\, dk. \tag{3.3}$$

where the value of σ_v can be predicted in renormalized perturbation theory (RPT) [148], although it is usually tuned to mock catalogues or N-body simulations. In order to minimize the systematic effect of incorrectly modelling non-linearities it is also important to perform the fit in a range of scales safely inside the linear or mildly non-linear regime ($r \gtrsim 30\,{\rm Mpc}/h$).

Now, two points at redshift z separated by a distance $r_\parallel$ along the line of sight have a redshift separation $\delta z = H(z)\, r_\parallel$. If the separation is completely transverse, they will subtend an angle $\theta = r_\perp / r_A(z)$, where $r_A(z) \equiv (1+z)\, d_A(z)$ is the comoving angular diameter distance to redshift z. Since we are working with the angle-averaged correlation function, we are not measuring any of these two, but a combination of both. This combination is usually modelled in BAO analyses as the geometric mean of the two transverse dimensions and the longitudinal one [133]:

$$r_0 = (r_\parallel\, r_\perp^2)^{1/3} = d_V(z)\, (\delta z\, \theta^2)^{1/3}, \tag{3.4}$$

where we have defined the volume averaged distance $d_V(z) \equiv \left[r_A^2(z)/H(z)\right]^{1/3}$. As we have seen, radial and transverse distances scale differently, and therefore, using an incorrect fiducial cosmology will introduce an anisotropy in the correlation function [149] (besides the one induced by RSDs). When using the monopole to measure the BAO this anisotropy is usually not considered, and the assumption is that the overall effect of using a wrong fiducial cosmology can be accounted for by scaling all distances with $d_V(z)$ [133]. Thus, measured scales are related to the real ones through

$$r_{\rm m}(z) = \frac{d_V^{\rm fid}(z)}{d_V(z)}\, r_{\rm real}. \tag{3.5}$$

Given that the template used in the fit has a BAO bump located at $r_s^{\rm fid}$, the real peak is at $r_{s,{\rm m}} = r_s^{\rm fid}/\alpha$, and therefore the parameter α constrains the ratio $d_V(z)/r_s$:

$$\frac{d_V(z)}{r_s} = \alpha\, \frac{d_V^{\rm fid}(z)}{r_s^{\rm fid}}. \tag{3.6}$$

To summarize, the method to measure the BAO from the monopole correlation function is:

1. Assuming a fiducial cosmology, translate redshifts into distances and calculate the monopole correlation function as explained in appendix A.2.2.

2. Fit the model in eq. (3.1) to the measured 2PCF, where $\xi_{\rm th}(r)$ is the theoretical monopole for the fiducial cosmology. This will yield a value for α and its error by marginalizing over the other parameters.

3. α constrains the ratio $d_V(z)/r_s$ (eq. (3.6)), which can then be used to obtain constraints on the cosmological parameters.

3.1.3 Statistical uncertainties and the covariance matrix

The best-fit value of α is found by performing a χ^2 analysis. For a measurement of the correlation function at N_r different values of r, the χ^2 is calculated as

$$\chi^2(\alpha) = \sum_{i,j=1}^{N_r} \left[\hat{\xi}(r_i) - \xi_{\rm fit}(r_i,\alpha)\right] C_{ij}^{-1} \left[\hat{\xi}(r_j) - \xi_{\rm fit}(r_j,\alpha)\right], \tag{3.7}$$

where C_{ij} is the covariance matrix of the correlation function. Even though the errors in the power-spectrum are, in principle, uncorrelated, this is not true for the 2PCF. Figure B.1 shows the reduced covariance matrix $\rho_{ij} \equiv C_{ij}/\sqrt{C_{ii}C_{jj}}$ for the angular (top-left panel), radial (top-right panel), monopole (bottom-left panel) and anisotropic (bottom-right panel) 2PCF. From this figure we can see that errors in the angular (transverse) direction are correlated over a much larger number of bins than in the radial direction. Since angular scales weigh a lot more than radial ones in the monopole (2 transverse dimensions versus 1 radial), the non-diagonal elements of the covariance cannot be ignored.

Estimating the covariance matrix is one of the most critical and complicated tasks in this process. There exist three main approaches to doing this:

- **From the data.** The survey is divided into N_R equivalent regions and they are treated as independent realizations of the galaxy distribution. The covariance matrix is calculated as

$$\hat{C}_{ij} = \frac{1}{N-1}\sum_{n=1}^{N_R} \left[\xi_n(r_i) - \bar{\xi}(r_i)\right] \left[\xi_n(r_j) - \bar{\xi}(r_j)\right], \tag{3.8}$$

  where ξ_n is the 2PCF calculated on the n-th region and $\bar{\xi} = \sum_n \xi_n/N_R$. The N_R different regions will have a much smaller volume than the whole survey, and therefore the cosmic variance will be different. For this reason, the *jackknife* method is preferred instead: the N_R different realizations are constructed by *omitting* each one of the regions from the whole survey. The caveat here is that the jackknife samples are not independent. The main advantage of this approach is that the errors are computed from the data in a model-independent way without extra assumptions. There exist other approaches to calculating the covariance from the data, such as using the 4-point function or the *Bootstrap* method (although the latter may be a bad idea for 2PCF studies - see chapter 3 of [150]).

- **From mocks.** Another possibility is to fabricate different realizations of the survey from computer simulations. The most precise way to do this would be to use an N-body simulation, however we have seen that they can be computationally expensive. Other approximate and faster methods have been proposed, such as the use of second-order Lagrangian perturbation theory [151], the Halo Model [152] or lognormal realizations [153, 154]. Once the matter density field is generated, a halo finder is run, galaxies are placed in the haloes (using for example a halo occupation distribution (HOD) scheme [101, 155, 156]) and the data are cut to the survey's field of view. From these realizations the covariance matrix is calculated as above. These methods have a few caveats: first, to simulate each realization one must use a fiducial cosmological model, and therefore the errors will not be model-independent. Secondly, one needs a large number of realizations in order to obtain a reliable estimate of C_{ij}, and generating the mocks may be

computationally costly. Finally, it is not always possible to simulate every observational effect, which may influence the magnitude of the errors.

- **From theory.** The covariance matrix can be estimated theoretically using Gaussian statistics for the overdensity field. These methods are described in appendix B. This is a very fast way of calculating C_{ij}, and can be done in a model-independent way. However many effects cannot be included in this formalism (including the fact that δ is actually non-Gaussian), and Gaussian estimates tend to be overly optimistic.

The covariance matrix receives two main contributions. First there is *c*osmic variance: the fact that we only have access to one realization of the Universe, and therefore our statistical uncertainties are limited by the volume of the region of the Universe we can see (we have assumed that we can apply the ergodic theorem - see appendix A.1). The overall amplitude of this contribution will be inversely proportional to the number of modes that fit inside the survey volume, and therefore will grow with V^{-1} (see eq. (B.5)). Then, for a galaxy survey there is an uncertainty related to the number of tracers of the matter density field we have used. This contribution is called *shot noise* and is a pure Poisson noise term that depends only on the observed number density. The contribution to the error of the 2PCF in a given bin is inversely proportional to the number of pairs found in that bin.

3.1.4 The effect of non-linearities

Since there exists no analytical method able to predict the power spectrum on all scales due to the non-linear nature of structure formation, it is reasonable to ask whether non-linear contributions alter the location of the BAO peak. If this is so and the effect is not correctly modelled, it could be a potential source of systematic uncertainties. Fortunately, at the scale of the BAO ($r_s \simeq 150\,\mathrm{Mpc}$) non-linear evolution is not so relevant. Non-linearities affect the BAO peak in two ways [157]:

- **Broadening the peak.** Non-linear peculiar motions broaden the peak, and the total peak width can be thought of as the quadrature sum of the linear width and the r.m.s. relative displacement. In the power spectrum this has the effect of damping the BAO wiggles on scales below $\sim 7-8\,\mathrm{Mpc}/h$. This broadening does not bias the estimation of the peak position as long as a suitable template is used in the fitting, but it may deteriorate its precision.

- **Shifting the peak.** Non-linear corrections can shift the peak position by moving pairs systematically closer or further away, however this effect is very small and only appears at second order in perturbation theory [158]. This shift has been studied both analytically and numerically [148, 159] and has been shown to have a very small effect on the BAO estimation, below $\lesssim 0.3\%$ at $z = 0$. Similar shifts may come from scale-dependent non-linear bias, and this contribution is comparable to the amount quoted above, depending on the bias model.

As we have seen, although non-linear effects are small, they may degrade the precision of the peak location and even introduce a small systematic. For this reason it was suggested in [160] that it might be possible to correct for these non-linear effects on the data themselves by reverting the non-linear gravitational evolution on each galaxy using the Zel'dovich approximation (or LPT in general). This technique is called *BAO reconstruction*, and has been studied analytically and numerically [159, 161], as well as used in surveys. In all cases it has been shown that by using reconstruction techniques the non-linear broadening and shifting of the peak almost disappears, and that the precision of the BAO improves significantly

(although this improvement is not so relevant in the latest 9th data release from the SDSS collaboration - see [30]).

3.2 Angular BAO in photometric redshift surveys

In order to improve the constraints on the cosmological parameters, several photometric redshift surveys have been planned, and are presently being carried out, such as DES [143], PanSTARRS [144], LSST [145] or PAU [146]. This improvement will be possible due to the larger volume and number of observed sources compared to previous spectroscopic surveys, even though photometric redshifts have lower precision compared to their spectroscopic counterparts. As an example, the CFHTLS survey estimated a photo-z error $\sigma_z \sim 0.03(1+z)$ for $0.2 < z < 1.5$ [162]. Such an uncertainty prevents a correct reconstruction of the radial information, and only allows in practice a study of the angular statistics of galaxy clustering, such as the angular 2PCF $w(\theta)$ or the angular power spectrum C_l.

This section is based upon the paper [1], in which an empirical parametrization of $w(\theta)$ is proposed as a means to recover the angular BAO scale as a function of redshift. It is important to stress the fact that our aim is not to use the full shape of the correlation function, but only to use the BAO as a standard ruler. Even though this reduces our sensitivity to the cosmological parameters, the position of the BAO turns out to be extremely robust against systematic uncertainties. It is unclear whether the full description would be more constraining, once all the systematic errors are correctly accounted for.

In spite of the different complications that measuring the angular BAO involves, such as projection effects (see section 3.2.3 below), this method is able to recover the input cosmology from the position of the BAO. This technique has been checked against theoretical calculations and has also been tested in a mock catalogue from an N-body simulation.

3.2.1 Modelling angular clustering

The galaxy-galaxy angular 2PCF $w(\theta)$ is the excess probability (over a random point distribution) of finding two galaxies subtending an angle θ

$$P(\theta)\, d\Omega_1\, d\Omega_2 = \bar{\sigma}_n^2\, (1 + w(\theta))\, d\Omega_1\, d\Omega_2, \tag{3.9}$$

where $\bar{\sigma}_n \equiv N_g/\Omega$ is the mean angular galaxy number density. As explained in appendix A.1.4, this is directly related to the 2PCF of the 2D projected overdensity

$$\delta(\hat{\mathbf{n}}) \equiv \int_0^\infty \delta(z,\, \hat{\mathbf{n}})\, \phi(z)\, dz. \tag{3.10}$$

The details about the theoretical modelling of $w(\theta)$ used in this work are described in appendix A.1.4, and will not be repeated here. The linear matter power spectra for different cosmologies were obtained using `CAMB` [88], and the effect of non-linearities on the BAO was included through the RPT Gaussian damping factor

$$P_{\rm NL}(k) = P_L(k)\, e^{-k^2\sigma_v^2/2}, \tag{3.11}$$

The covariance matrix of $w(\theta)$ was estimated using the Gaussian prediction given by [163]

$$\mathrm{Cov}(\theta_1, \theta_2) = \frac{2}{4\pi\, f_{\rm sky}} \sum_{l=0}^{\infty} \frac{2l+1}{4\pi} L_l(\cos\theta_1)\, L_l(\cos\theta_2) \left(C_l + \frac{1}{\bar{\sigma}_n}\right)^2, \tag{3.12}$$

where $f_{\rm sky}$ is the fraction of the sky covered by the survey. We verified that this estimation is compatible with the jackknife errors computed from the simulation. Further details about estimating Gaussian covariance matrices can be found in appendix B.

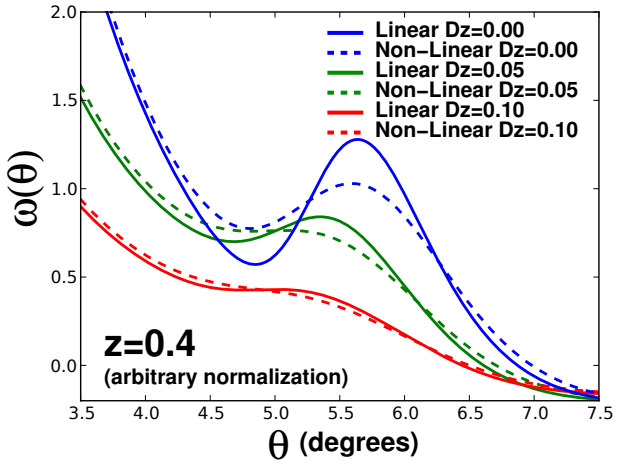

Figure 3.1: Illustrative representation of the effect of non-linearities and finite bin widths on the angular correlation function. The different colours correspond to different values for the width Δz, and solid and dashed lines represent the linear and RPT predictions respectively. Both non-linearities and the projection of the galaxy density in a finite redshift bin widen the shape of the peak. The projection also displaces the position of the peak to smaller scales. This effect increases with the bin width, and is larger at smaller redshifts.

3.2.2 The angular sound horizon scale

The robustness of BAO as a standard ruler lies in the ability to relate the position of the BAO bump in the 2PCF to the sound horizon at decoupling r_s, however these two quantities are not exactly equal [164, 165, 158, 159, 166]. It is useful to distinguish the following two angular scales: the angle subtended by the sound horizon $\theta_{\rm BAO}(z) \equiv r_s/\chi(z)$ and the location of the BAO peak, defined as the position of the local maximum in the linear angular correlation function using an infinitesimal redshift bin $\theta_{\rm p}$. These two scales are offset by about 1 or 2 percent [165], depending on the cosmological model, but there are other effects to take into consideration.

First, non-linearities damp the BAO wiggles and thus widen the peak, decreasing the significance of its detection as a local maximum. Furthermore, in practice it is not possible to use infinitesimal redshift shells, due to sampling variance, shot noise and also, for a photometric survey, due to the uncertainty in z. Using finite redshift bins entails dealing with projection effects which, as we will see shortly, alter the shape and position of the BAO. Finally, redshift-space distortions have an impact on $w(\theta)$ and may contribute to this effect [167].

For a photometric redshift survey, wide redshift bins are needed. Implicitly we will be averaging over angular correlation functions at different redshifts within our bin, which have different BAO positions. This induces two projection effects:

- The peak position shifts towards smaller angles, and the displacement is larger for larger bin widths.

- The amplitude of the peak is reduced until the local maximum disappears, leaving only a "shoulder" [168, 169]. As an order of magnitude we have computed that there will be a well defined local maximum for $\bar{z} > 0.6$ for a photometric redshift error $\sigma_z \simeq 0.01\,(1+z)$, but the maximum can only be found at $\bar{z} > 1.3$ for $\sigma_z \simeq 0.03\,(1+z)$.

These effects are illustrated in figure 3.1.

In [1] we proposed a method to measure the BAO scale alternative to the location of the local maximum. Currently, the ratio of the power spectrum to a smooth fit is used, to avoid the effect of the tilting function, which changes the peak position [148, 159, 166, 170]. The method described in [1] is based on the angular correlation function in configuration space, and will be the subject of the next sections.

3.2.3 Method to recover $\theta_{\rm BAO}$

Our method is based on an empirical parametrization of $w(\theta)$. The full recipe to estimate $\theta_{\rm BAO}$ is:

- Divide the full galaxy sample into sufficiently wide redshift bins.
- Compute the angular 2PCF in each bin as explained in section A.2.2.
- Fit the following 6-parameter model to the data:

$$w(\theta) = A + B\,\theta^{\gamma} + C\,e^{-(\theta-\theta_{\rm FIT})^2/2\sigma^2}. \tag{3.13}$$

- The value of $\theta_{\rm BAO}$ is obtained from the parameter $\theta_{\rm FIT}$ correcting it for the projection effects we have described. Since our power law + Gaussian parametrization is an empirical model, only valid in the neighbourhood of the peak, the other parameters have little physical interpretation. We use the amplitude of the Gaussian term C and its error to determine the significance of the BAO detection.
- Fit the cosmological parameters to the evolution of $\theta_{\rm BAO}$.

Parametrization of $w(\theta)$

The correlation function $w(\theta)$ is parametrized as in equation (3.13). Our model uses a power law to describe the broadband shape of the correlation function before and after the peak, which is described by the Gaussian. The mean $\theta_{\rm FIT}$ of the Gaussian term is our proxy for the BAO scale, and the parameters B and C describe the relative weights of the two contributions.

We have tested this parametric model in a redshift interval from 0.2 to 1.4 for different redshift bin widths (from 0 to 0.2) and for the 14 different cosmological models listed in table 3.1. The parametrization is good to 1% (i.e. the values of the χ^2/ndof are between 0.98 and 1.01 and the corresponding probabilities between 0.6 and 0.9) for errors of $\sim 1\%$ for all angles. This precision is much better than that expected for any realistic redshift survey. A few examples of these fits are shown in figure 3.2. Other functional forms have been tried, but none showed any clear advantage over our final choice.

As we have said, our empirical parametrization is only intended to work within a range of angular scales around the BAO peak. The fitting range must therefore be restricted, since our model cannot accommodate non-linear scales, and the correlation function changes its slope after the BAO. There is a region of stability in which the start and end points of the fit can be safely chosen (more in section 3.2.4).

Correcting for projection effects

The shift in the peak position due to the projection effect must also be dealt with in our method. Our ansatz for this correction is

$$\theta_{\rm BAO} = \alpha\,\theta_{\rm FIT}, \tag{3.14}$$

h	Ω_M	Ω_b	Ω_k	w_0	w_a	n_s
0.70	0.25	0.044	0.00	-1.00	0.0	0.95
0.68						
0.72						
	0.20					
	0.30					
		0.040				
		0.048				
			+0.01			
			-0.01			
				-0.90		
				-1.10		
					-0.1	
					+0.1	
						1.00

Table 3.1: Summary of the 14 cosmological models used to test the method. Where empty, the fiducial values (first line) are assumed.

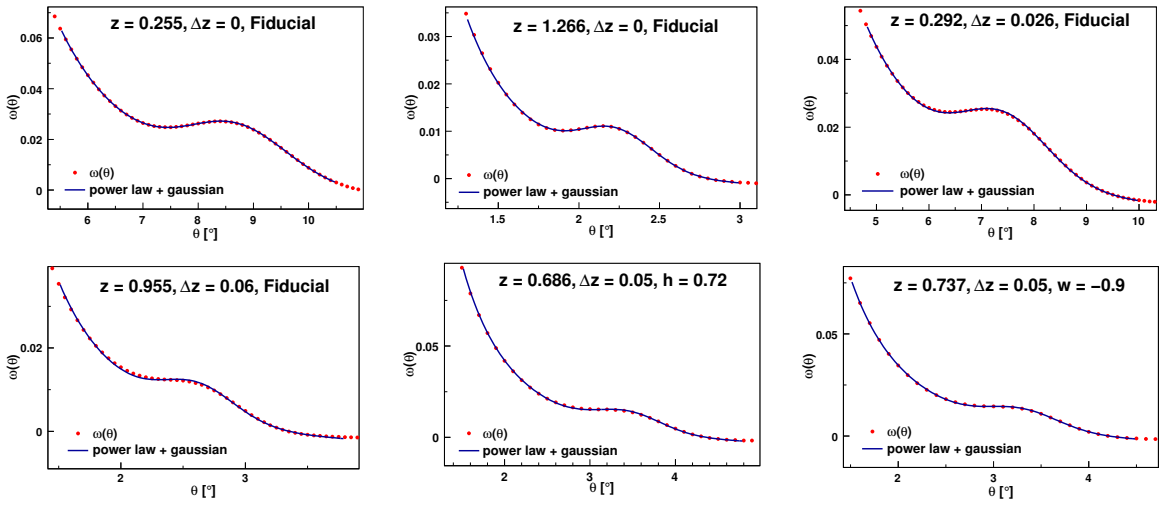

Figure 3.2: Goodness of the parametrization for several redshifts, bin widths and cosmological models.

where α could in principle be a function of redshift, bin width and cosmology. However, in order to obtain unbiased cosmological constraints, α must be cosmology-independent. We have tested this for the 14 different cosmologies summarized in table 3.1 and for different bin widths.

This was done by producing theoretical correlation functions for each case, modelled as described in section 3.2.1. This modelling includes the survey redshift distribution (eq. (3.15)). We then apply our fit to the theoretical correlation functions and compute the projection factor α. The two main results from this study are:

- The shift has a universal form, virtually independent of the cosmology. Thus the correction depends only on redshift and bin width $\alpha = \alpha(z, \Delta z)$ within a precision $< 1\%$.
- After applying this correction we manage to recover the true BAO scale for all the cosmological models to a precision better than 0.75%. In particular, this correction is able to account for the 2% offset due to the tilting function observed in [165].

These two results are illustrated in figure 3.3, where we plot α for different redshifts, bin widths and for the 14 cosmological models. As is shown, the spread due to the different cosmological parameters is much smaller than the correction itself.

As has already been noted [169], the projection effect is much more pronounced at low redshift, a fact that must be taken into account when choosing the optimal bin width. The correction must be kept as small as the photo-z errors allow it, in order to minimize systematic uncertainties. In our analysis we use a Gaussian photo-z error, since if the requirements on photo-z errors for a survey like DES are fulfilled, the effects of a possible non-Gaussianity will be small. If this is not the case, the evaluation of the true bin width must be refined, which is another reason to maintain the correction small. The effect of non-Gaussian photo-z errors is twofold. First they increase the final error due to the influence of the photo-z on $\theta_{\rm BAO}$, and secondly, they increase the correlation between redshift bins. A small fraction of the galaxy sample ($\mathcal{O}(1\%)$) with large photo-z errors (2-3 times larger than the nominal value) could translate into an increase of $\sim 10\%$ in the final error in $\theta_{\rm BAO}$.

Note that the BAO scale is correctly recovered for both linear and non-linear theory for infinitesimal bins. This is shown in figure 3.4, where the residual with respect to the theoretical $\theta_{\rm BAO}$ as a function of redshift is shown for the fiducial cosmological model. The recovered values are well inside the 0.75 per cent precision that we quote as systematic error, represented by the dotted lines. This happens for all the cosmological models, showing that the method is robust and able to recover the theoretical BAO scale.

Redshift-space distortions

RSDs must also be taken into account in the parametric description of the angular correlation function. Galaxy pairs separated by large radial distances infall towards each other, and their measured relative radial distance is smaller than the real one. This results in larger correlations at intermediate scales ($\pi < 20\,{\rm Mpc}/h$) and lower ones at larger separations. The radial correlation function becomes negative around $\pi \sim 40\,{\rm Mpc}/h$ (see [171]). For narrow bins, the angular 2PCF in real and redshift space can be quite different, however, for large bins this difference decreases, since the total number of pairs is conserved (see figure 3.6).

Figure 3.5 shows the linear theory prediction for the angular correlation function with and without RSDs and photo-z, as well as the value of $\theta_{\rm FIT}$ from our parametrization. There is a small but systematic shift of $\theta_{\rm FIT}$ when including the photo-z. This is due to the fact that the real bin width is larger in this case, and the systematic offset is corrected by α, using the

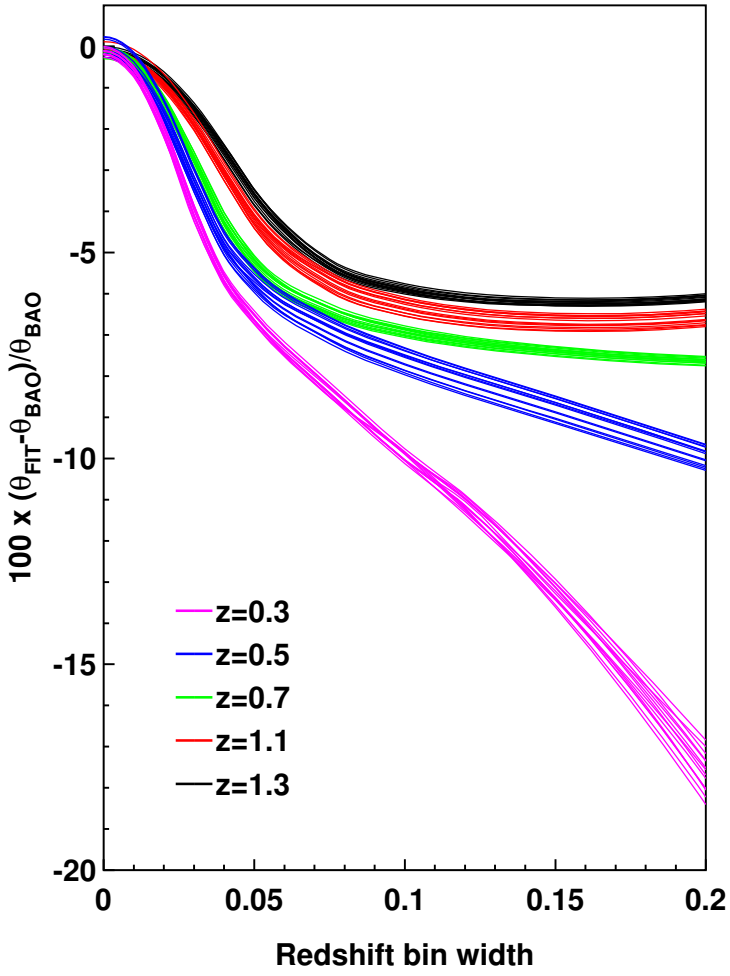

Figure 3.3: Evolution of the shift $\theta_{\rm FIT} - \theta_{\rm BAO}$ with the width of the redshift bin for the 14 cosmological models considered in this work. The sound horizon scale $\theta_{\rm BAO}$ for each cosmological model is taken as the reference. Two results are obtained. First, the correct sound horizon scale is recovered for any cosmological model within 0.75% for infinitesimal redshift shells. Second, the shape of the shift is universal, and the spread is constant with the redshift bin width. Note that the dispersion between models is much smaller ($\leq 1\%$) than the shift due to increasing bin width. At low redshifts, $z < 0.3$, the shift can be $\sim 15\%$ for $\Delta z \geq 0.1$, while at high redshifts, $z > 0.5$, the shift saturates at around 7%.

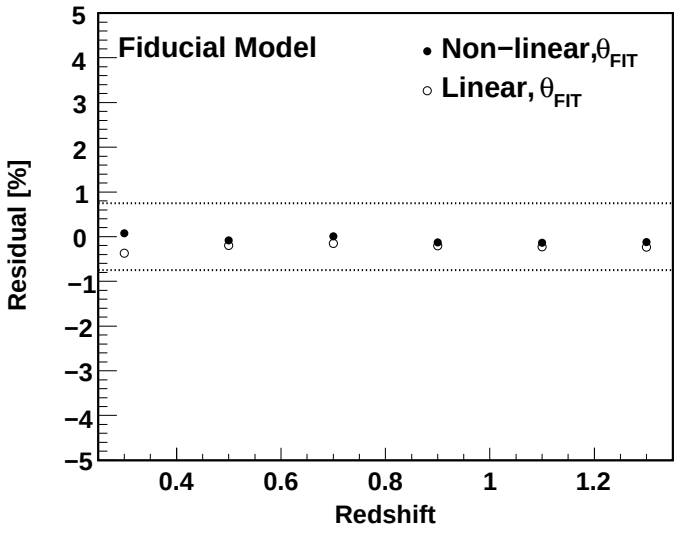

Figure 3.4: Residuals of the result of the parametric fit $\theta_{\rm FIT}$ with respect to the the theoretical value of the BAO scale as a function of redshift for the fiducial model. The open dots show the results obtained from the fit to the linear correlation functions and the solid dots show the results on the non-linear correlation functions. In both cases the theoretical values are recovered, well inside the quoted systematic error of 0.75%, shown by the dotted lines.

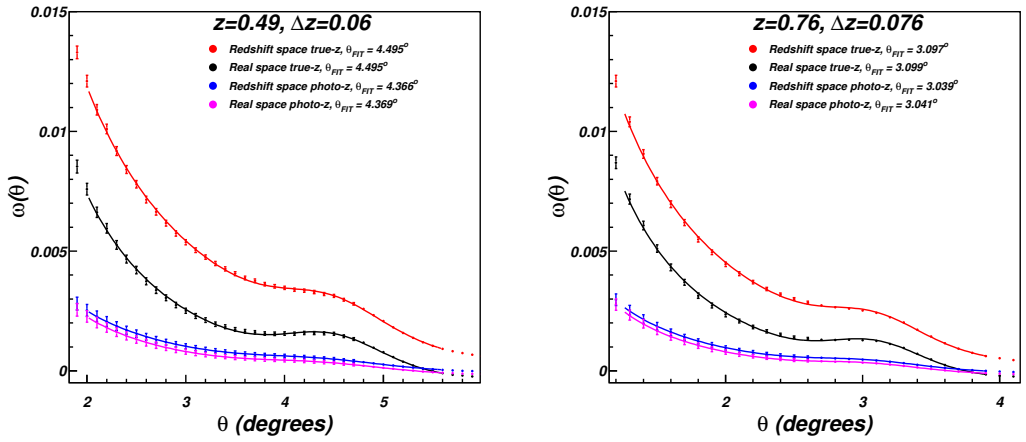

Figure 3.5: Effect of RSDs and photo-z. Points (with nominal 2% errors) show the linear theory prediction, while lines correspond to a fit to our parametrization. Labels indicate the corresponding values of $\theta_{\rm FIT}$ in each case.

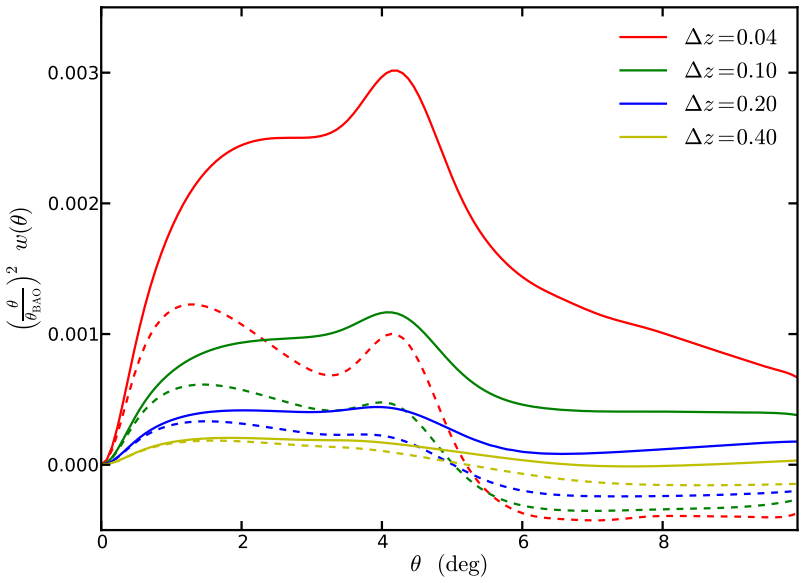

Figure 3.6: Angular correlation function in redshift space (solid lines) and real space (dashed lines) for different bin widths. By projecting the galaxy density on wider bins the RSDs gradually disappear, and both correlation functions are more similar.

corresponding bin width in true redshift. This figure also shows that our method is able to absorb the effect of RSDs despite the change in the amplitude of the 2PCF.

3.2.4 Applicability to the Dark Energy Survey

The Dark Energy Survey

DES is a sky survey aimed specifically at understanding the nature of the dark energy. It will use 30 per cent of the available time on the Blanco telescope at Cerro Tololo Inter-American Observatory, for 5 years. The survey was designed to measure photometric redshifts in the range $0.2 < z < 1.4$ and will cover $5000\,\mathrm{deg}^2$ in the Southern hemisphere. The filters to be used are g, r, i, z and Y. DES started its operations in November 2012 and is currently in the science validation stage [172].

DES has been designed to exploit mainly four methods in order to study the dark energy: supernovae (~ 2000 expected), weak gravitational lensing, galaxy cluster statistics and galaxy clustering ($\mathcal{O}(10^8)$ galaxies with sufficiently good photo-z's). The forecasted constraints on the equation of state of dark energy are $\Delta w/w \sim 5-15\%$ and about 30% in its variation with time w_a.

N-body simulation

In order to test our method in a more realistic case, we have applied it to a simulated galaxy catalogue. The catalogue was provided by the MICE[1] project team and was produced from a large N-body simulation specifically designed to reproduce the area, number density, depth

[1] http://maia.ice.cat/mice/

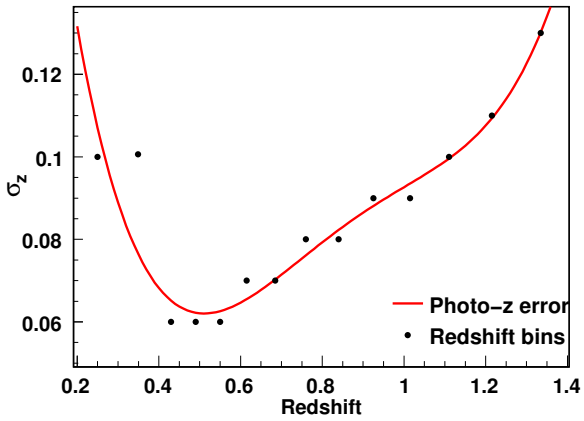

Figure 3.7: Photo-z precision used in this analysis (continuous line), compared to the bin position and widths used (dots).

and redshift distribution of DES. In particular the radial selection function used was

$$\frac{dN}{dz} \propto \left(\frac{z}{0.5}\right)^2 \exp\left[-\left(\frac{z}{0.5}\right)^{1.5}\right]. \tag{3.15}$$

The catalogue was extracted from the comoving output of an N-body simulation at $z = 0$. The box size of this simulation is $3072\,\mathrm{Mpc}/h$ and it contains 2048^3 particles, with a mass resolution $m_p = 2.3 \times 10^{11}\, M_\odot/h$. The cosmological model corresponds to the fiducial one in table 3.1: $(\Omega_\mathrm{M}, \Omega_\Lambda, \Omega_\mathrm{b}, h, w, \sigma_8, n_s) = (0.25, 0.75, 0.044, 0.7, -1, 0.8, 0.96)$. More details about this run can be found in [173, 174].

The mock galaxies were given a photometric redshift, computed from their true redshift by adding a random Gaussian contribution with a width corresponding to the desired photo-z accuracy. Figure 3.7 shows the photo-z variance used for this catalogue as a function of redshift, and corresponds to the typical precision obtained by using the current photo-z codes for DES [175]. Figure 3.7 also shows the redshift bins used in this analysis. These are shown as dots: their position in the horizontal axis corresponds to the mean redshift, while the vertical coordinate corresponds to the bin width. The bin widths were chosen as a compromise between statistics, photo-z precision and minimizing projection effects. These widths follow the expected photo-z precision, being wider at low and high redshifts. Wider bins affect the sensitivity of the measurement, since the amplitude of the correlation decreases and the shift in the peak position becomes very large, and narrow bins are very correlated, complicating the analysis.

Results

The correlation functions were obtained from the data using the Landy and Szalay estimator [176] (this is described further in appendix A.2.2). The random catalogues used contained the same number of galaxies as the data, since shot noise is negligible given the size of the sample. The 14 correlation functions, together with their parametrizations are shown in figure 3.8. The statistical errors shown were calculated as in equation (3.19), taking into account the photo-z effect as we will describe shortly.

The recovered values of θ_BAO after applying the correction (eq. (3.14)) can be seen in fig. 3.9 (top panel) as a function of redshift. The correction is applied after evaluating the true

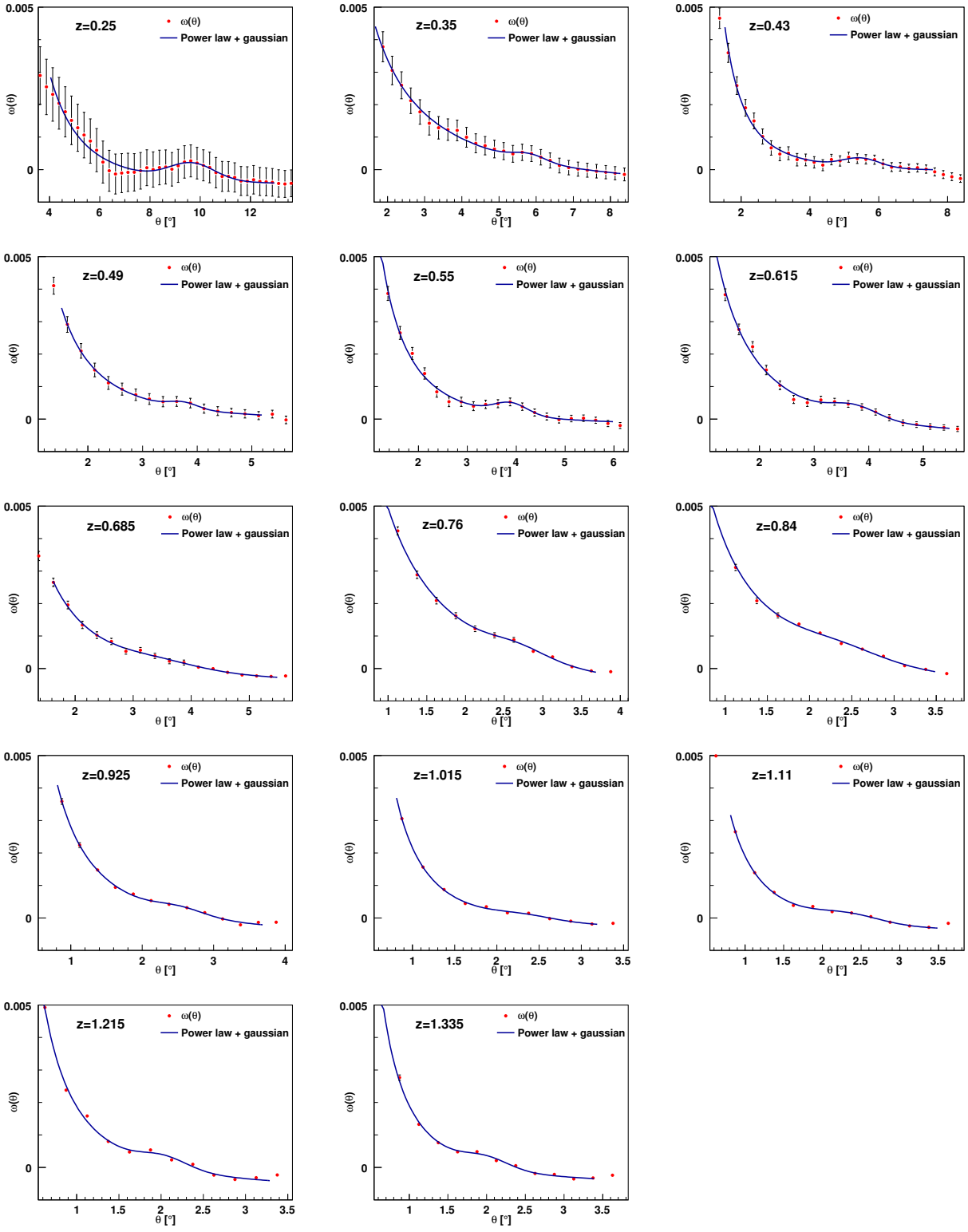

Figure 3.8: Correlation functions for MICE simulation, including photo-z effects.

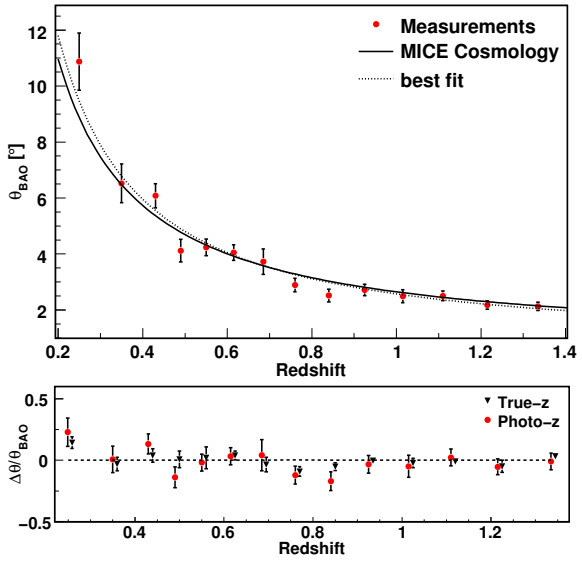

Figure 3.9: Evolution of the measured $\theta_{\rm BAO}$ with redshift (top). Comparison of the results obtained using photo-z and true-z (bottom). The true-z results are closer to the MICE cosmology, as expected.

width of the redshift bin:

$$\Delta z_{\rm true} = \sqrt{2\pi}\Delta z_{\rm photo}, \tag{3.16}$$

which corresponds to the width of a top-hat bin in true-z giving the same amplitude in $w(\theta)$ as the photo-z bin. The points shown also include systematic errors, described below. The cosmology for the MICE simulation is shown as a solid line and the best-fitting cosmology as a dashed line. The same analysis has been carried out using the true redshifts, and the results are presented in the bottom panel of figure 3.9 together with those of the photo-z. The results for true redshifts are closer to the fiducial cosmology, as expected. In this case the bin width is not corrected, and the systematic errors do not include any photo-z contribution.

After applying our method the true cosmology is recovered, both for the photo-z and the true-z results. The constraints obtained with the former are shown in figure 3.10, where the dot inside the 1-*sigma* region corresponds to the fiducial cosmology. The contours were obtained by fitting the cosmology to the values of $\theta_{\rm BAO}$ using $\Omega_{\rm M}$ and w as free parameters and fixing all the other parameters to their fiducial values.

In order to perform this fit, it is necessary to take into account the non-zero correlations between the values of $\theta_{\rm BAO}$ measured in different bins, caused by the photo-z. We estimate this correlation by computing the migration matrix r_{ij}, defined as the probability that a given galaxy at bin i is measured at bin j due to the photo-z error. Thus the number of galaxies observed inside the i-th bin is

$$N_i^{\rm O} = \sum_{j=1}^{N_b} r_{ij}\, N_j^{\rm T}, \tag{3.17}$$

where $N_i^{\rm T}$ is the true number of galaxies at redshift i. The observed and true correlation functions are related by [177]

$$w_i^{\rm O}(\theta) = \sum_{j=1}^{N_b} r_{ij}^2 \left(\frac{N_j^{\rm T}}{N_i^{\rm O}}\right)^2 w_j^{\rm T}(\theta). \tag{3.18}$$

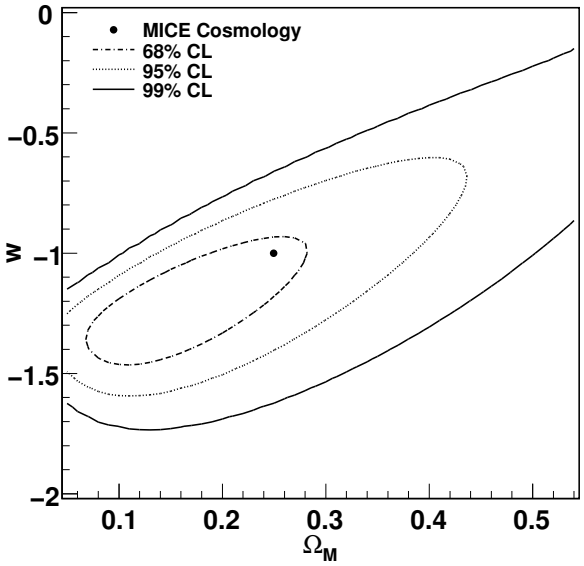

Figure 3.10: Contours for $\Omega_{\rm M}$ and w at 68 (dashed-dotted line), 95 (dashed line) and 99 (solid line) % confidence level. The true cosmology, given by the dot, is recovered within 1-σ.

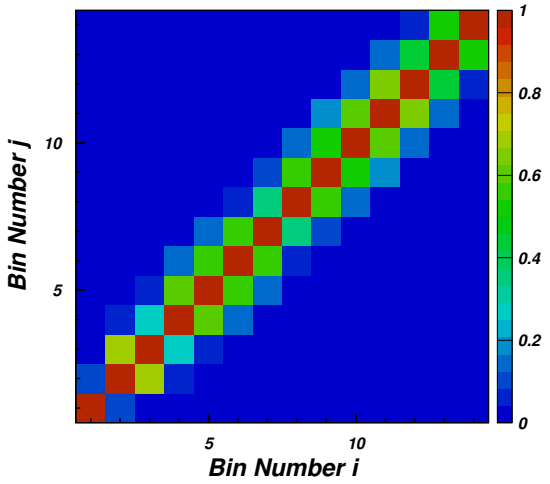

Figure 3.11: Correlation matrix for redshift bins. The correlation extends to 3 bins, and is taken into account in the fit of the cosmological parameters.

Systematic error	$\Delta\theta_{BAO}$	Correlated between bins
Parametrization	1.0%	No
Photometric redshift	5.0%	Yes
Redshift-space distortions	1.0%	Yes
Theory	1.0%	No
Projection effect	1.0%	No

Table 3.2: Estimation of various systematic errors. Some of them are correlated between redshift bins (photo-z and RSD errors).

Hence, the covariance for the correlation functions observed in different bins is

$$C_{ij}(\theta_1,\theta_2) = \sum_{k=1}^{N_b} r_{ik}^2 r_{jk}^2 \left(\frac{(N_k^{\mathrm{T}})^2}{N_i^{\mathrm{O}}\, N_j^{\mathrm{O}}}\right)^2 \mathrm{Cov}(\theta_1,\theta_2), \tag{3.19}$$

where $\mathrm{Cov}(\theta_1,\theta_2)$ is given in equation (3.12). This expression neglects the intrinsic correlation between bins which would be present even if true redshifts were used. Whether this is a good approximation will depend mainly on the bin widths. In our case, bin centres are separated by more than 200 Mpc/h, and it should be safe to ignore these intrinsic correlations. Thus, with photo-z errors, the observed 2PCFs and covariances are given by equations (3.18) and (3.19). The correlation coefficients between different redshift bins computed form C_{ij} are shown in figure 3.11. We see non-negligible correlations extending up to three bins.

The covariance matrix for the correlation functions is transformed into the covariance of θ_{BAO}, $\hat{C}_{ij}$, and the χ^2 of the fit to a given cosmology is

$$\chi^2(\Omega) \equiv (\theta^i_{\mathrm{BAO}} - \theta_{\mathrm{BAO}}(i,\Omega))\, \hat{C}^{-1}_{ij} (\theta^j_{\mathrm{BAO}} - \theta_{\mathrm{BAO}}(j,\Omega)), \tag{3.20}$$

where $\theta_{\mathrm{BAO}}(i,\Omega)$ is the theoretical prediction for the cosmology Ω in the i-th bin. The total error contains the systematic uncertainties which we will discuss shortly. Some of these are correlated (such as RSDs and photo-z),while others are just added in quadrature to the diagonal of the covariance matrix. These quantities satisfy the DES requirements: the total area of the survey and its galaxy number density (fixed by the survey magnitude limits) contribute to the total error on $w(\theta)$, and the photo-z requirements fix the correlations between bins.

Restricting ourselves to a 1-dimensional analysis and fixing Ω_{M} to its fiducial value, we obtain constraints of $\sim 15\%$ on the equation of state of dark energy $w = -1.05 \pm 0.14$. This result depends strongly on the quality of the photo-z, since it is by far the dominant source of systematic errors. For example, for a more optimistic and ideal estimation of $\sigma_z = 0.03\,(1+z)$, the error on w improves to $\Delta w \sim 0.1$. This is in good agreement with the Fisher matrix forecasts for BAO with DES [143].

Systematic errors

In the analysis we have studied different sources of systematic error listed below. The contribution ascribed to each of them is shown in table 3.2:

- **The fitting procedure (method error).** This error could be due to the influence of the limits within which the fit is performed. To assess the importance of this source, we have performed the fit on the theoretical correlation functions for the 14 redshift bins, varying the fit limits. We choose a starting point well before the BAO feature, and one

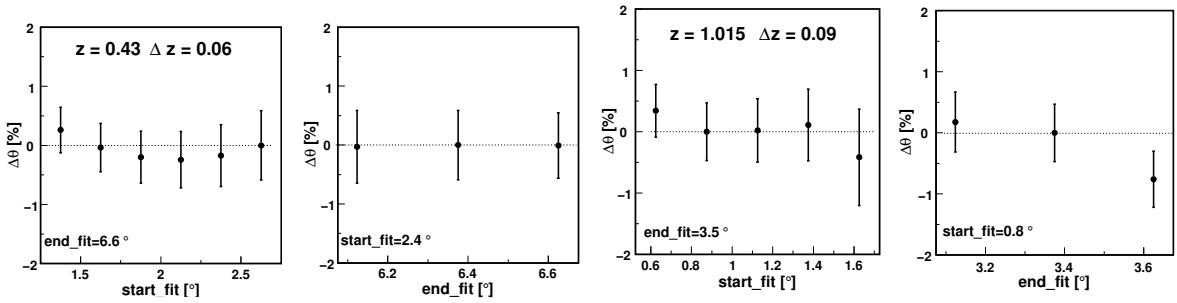

Figure 3.12: $\theta_{\rm BAO}$ dispersion as a function of the starting and end points of the fitted region. Only two bins are shown, where the maximum of the effect is observed, at redshifts 0.43 (top) and 1.015 (bottom). All the bins behave in a very similar way, but with smaller variations. The observed variation in the results is slightly smaller than 1%, and a conservative 1% systematic error is associated to the fit method.

after the peak, and see how the value of the $\theta_{\rm FIT}$ parameter varies in each case. The results for different limits are shown in figure 3.12 for the most extreme cases. In all cases the uncertainty is below the percent level, so we have conservatively assigned a value of 1% for this source.

- **Cosmology independence of the method.** This is caused by the residual dependence on the cosmological model of the correction factor $\alpha(z, \Delta z)$ for projection effects. The dispersion of the cosmological models around a common value for the projection effect correction is shown in figure 3.3 to be below 1%.

- **Theoretical modelling of non-linearities.** This error has been estimated in a conservative way as the maximum difference between the $\theta_{\rm FIT}$ measured from the theoretical linear 2PCFs and the ones containing non-linearities via the RPT damping factor. In all cases this effect is below 1%.

- **The photo-z.** To compute the systematic uncertainty associated to the error in the photometric determination of the redshift we have redone the analysis using the true redshifts of the mock catalogue. Figure 3.13 shows the distribution of the difference between the values of $\theta_{\rm FIT}$ found in both catalogues. As is shown, the photometric redshifts do not introduce a significant bias, but they cause a large dispersion of $\sim 5\%$ around the central value. The bottom panel of figure 3.13 shows the redshift dependence of the differences in $\theta_{\rm FIT}$, justifying our choice to assign a redshift-independent value of 5% to this source. Whether this assumption is true could be tested by using several realizations of the mock catalogue, and this study would be very relevant for a real survey, since this is the most important source of systematic uncertainty. However this is beyond the scope of this work.

- **Other sources.** We have identified other possible marginal sources of systematic errors: RSDs, bias and magnification. We have tested the possible shift in the position of the peak due to redshift-space distortions by varying the β parameter within its uncertainties, finding that this shift is below 1%.

  On the other hand, it is a good approximation to consider that on large scales (like the BAO scale), galaxy bias is scale-independent [178, 163]. The value of this bias may change in different redshift bins, but in this case $b(z)$ will act mainly as multiplicative factor, which our parametric fit can easily absorb without changing $\theta_{\rm FIT}$. We have also

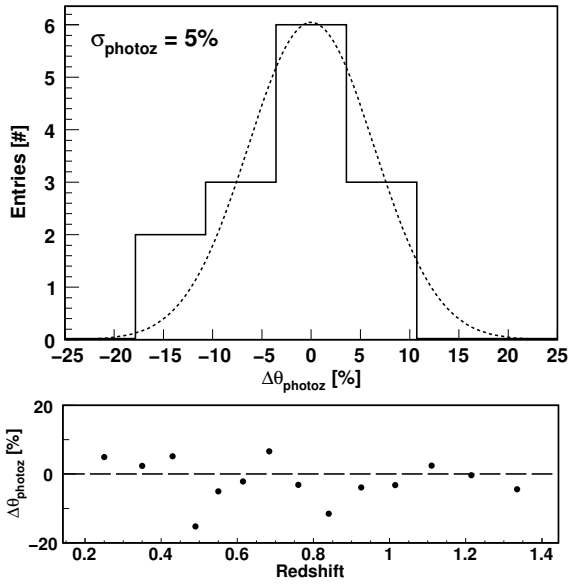

Figure 3.13: Cumulative Difference between the measured $\theta_{\rm FIT}$ in the photo-z catalogue and the true-z catalogue for the 14 bins of the analysis (top). The dispersion is 5%, that we use as the systematic error. This error is taken as constant with redshift, since there is no clear dependence when the differences are plotted as a function of redshift (bottom).

tested the effect of a scale-dependent bias by introducing it artificially in our correlation functions using an approximated Q-model [178]. The variation in the determination of the sound horizon scale in this case is still below 1%.

Finally, the gravitational lensing magnification can affect the angular correlation function [168] introducing correlations between different redshift bins. However, magnification acts mainly on small scales, and this correlation is much smaller than the one due to the photo-z. Therefore we did not consider this source.

It is important to compare the relevance of the different contributions to the total error at different redshifts. Statistical errors (essentially cosmic variance) dominate at small redshifts, where the available volume is smaller. The large number of observed objects due to the photometric redshifts prevents shot noise from taking over at higher redshifts, however, the systematic error associated with the photo-z becomes more important for $z < 0.5$. All other sources of systematic error remain subdominant for all z.

3.2.5 Discussion

The angular scale of the sound horizon can be recovered from redshift surveys (both spectroscopic and photometric). The method proposed in [1] is able to do it through an empirical parametric fit to the correlation function consisting of a power law term describing the background tilt and a Gaussian term to account for the BAO feature. As we have discussed, this parametrization is able to accommodate all the different effects that would be relevant for a photometric redshift survey, including RSDs, galaxy bias, non-linear growth and projection effects. Once the fit is performed, projection effects can be corrected for in a cosmology-independent way, and it is possible to recover the true BAO scale for a wide range of redshifts. This scale may now be used as a standard ruler to constrain cosmological parameters. Since

the method uses pure observables (angular positions for each redshift bin), its results are model-independent.

The method has been tested on a mock catalogue produced from an N-body simulation provided by the MICE team, designed to mimic the main characteristics of DES. We are able to recover the true cosmology of the simulation within 1-σ. The analysis carried out to obtain this results includes the correlations between redshift bins caused by the photo-z, as well as all the systematic contributions to the total error. Of these, the dominant one is by far the photometric redshift precision.

It must be noted that the analysis done in this work was performed in real space, rather than redshift space, since the mock catalogue did not include RSDs (nonetheless, the catalogue used for the final comparison with the conventional method, described in section 3.4, did contain RSDs). Also, dark matter particles were used as galaxies, and therefore no galaxy bias was simulated. As has been discussed, we believe that these are not essential limitations. If anything, both these effects would enhance the amplitude of correlations, and that may increase the sensitivity of our method.

To finalize, we would like to remark that this method has already been used on real data, yielding a measurement of the angular BAO from the photometric sample of LRGs from the SDSS DR7 [29].

3.3 Measuring the radial BAO scale

As has been reiterated along this chapter, the measurement of the BAO scale the galaxy power spectrum as a function of redshift is one of the most robust cosmological probes, since it is almost insensitive to systematic uncertainties related to the astrophysical properties of galaxies. Moreover, for recent and future data, with high quality and good statistics, it will be possible to extract information about the expansion history of the Universe from two different BAO observables: the angular diameter distance, through the measurement of the angular BAO scale, and the Hubble parameter, through the measurement of the BAO scale in the distribution of galaxies along the line of sight.

As we described in section 3.1, up to now most of the effort has been devoted to measuring the BAO signature in the angle-averaged correlation function, since it makes the most efficient use of the data. However, some groups have tried to measure the BAO scale in the purely radial direction [171, 179, 180]. In this section we describe in detail the results presented in the paper [2]. This time we propose a new method to extract the evolution of the radial BAO scale with redshift, and explain how to use it as a standard ruler to determine cosmological parameters. As was proposed for the angular case, we use data from galaxy surveys in a manner that is fully cosmology-independent, since only purely observable quantities are used in the analysis. The estimation of the radial BAO scale is based on a similar idea too: an empirical parametric fit to the radial correlation function. For this kind of measurement a good determination of the redshift is imperative, therefore these results hold only for spectroscopic surveys.

The method is designed to be used as a strict standard ruler, and provides the radial BAO scale as a function of redshift. Again, we do not intend to give a full description of the radial correlation function, since we consider this approach to be more robust against systematic effects. As we demonstrate, the measurement is only limited by cosmic variance, and the associated systematic errors are clearly subdominant.

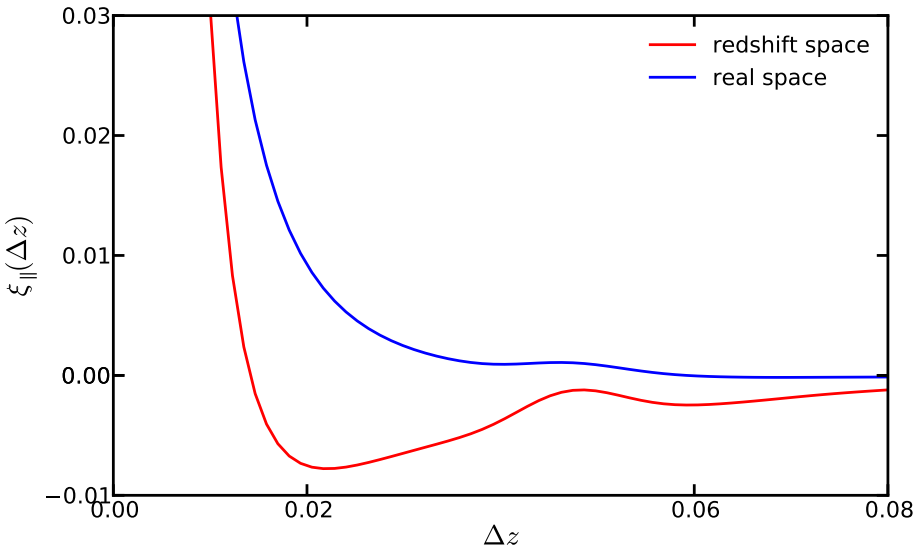

Figure 3.14: Radial correlation function in real space and redshift space. RSDs change the shape of the correlation function noticeably, creating negative correlations before the BAO (the dipole $\xi_2(r)$ is responsible for this - see figure 2.2).

3.3.1 Galaxy clustering in the radial direction

In the previous section we were interested in measuring the BAO scale in the angular or transverse direction. For this, bins were made in redshift so that mostly transverse galaxy pairs contributed to the correlation function. Our aim in this section is to measure the BAO feature in the radial correlation function, for which we must use pairs of galaxies whose relative separation is parallel to the line of sight. Of course, the number of perfectly aligned pairs is very small, and we must allow for some angular separation between pairs, in order to suppress both shot noise and and cosmic variance. Thus we will be studying the correlation function of galaxy number density projected onto the radial direction in angular pixels of some resolution $\delta\Omega^2$

$$\delta_{\parallel}(z) \equiv \frac{1}{\delta\Omega^2}\int_{\delta\Omega^2} d\theta\, d\phi\, \delta(t(z), r(z), \theta, \phi). \tag{3.21}$$

This is equivalent to the projection onto the transverse plane used for the angular case (eq. (3.10)).

For sufficiently good angular resolution, the correlation function of the projected field $\delta_{\parallel}$ is equivalent to the radial correlation function, defined as the 3D anisotropic correlation function with the transverse dimension σ set to 0 (or $\mu = 1$). The radial correlation function shows a BAO feature at $\pi \simeq r_s$, corresponding to galaxies whose redshift separation is

$$\pi_{\rm BAO} \equiv r_s = \chi(\bar{z} + \Delta z_{\rm BAO}) - \chi(\bar{z}) \simeq \frac{\Delta z_{\rm BAO}}{H(\bar{z})}. \tag{3.22}$$

Assuming that the evolution of the density field across the range of redshift explored can be ignored, and for an expansion rate that is slowly varying (as is the case for the standard cosmological model), the radial correlation function can be thought of as depending only on the redshift separation:

$$\xi_{\parallel}(\bar{z}, \Delta z) \simeq \xi\left(\pi = \frac{\Delta z}{H(\bar{z})}, \sigma = 0\right). \tag{3.23}$$

This is the model for the theoretical radial correlation functions that we used for this analysis. In order to calculate the 3D correlation function $\xi(r,\mu)$ the linear matter power spectra for the 14 cosmologies listed in table 3.1 were produced with `CAMB` [88]. As we did for the angular case, non-linearities were introduced through the RPT Gaussian damping factor (eq. (3.3)). Redshift-space distortions, which change the shape of the radial correlation function very noticeably (see figure 3.14), were included in the linear and plane-parallel approximation (i.e. through the Kaiser effect [78]). Hence the form of the anisotropic power spectrum used is

$$P^s_{\rm NL}(k,\mu_k,z) = (1+\beta(z)\,\mu_k^2)^2\,P^r_{\rm L}(k,z)\,e^{-k^2\,\sigma_v^2(z)/2}. \tag{3.24}$$

We expect this theoretical model to be correct for large scales, where the neglected effects (fingers of god, mode-mode coupling, scale-dependent bias) are very small. We have used a bias parameter $b=1$ throughout the calculation. The possible influence of bias is studied in the section on systematic errors.

Projection effects

As we have pointed out, the correlation function of the projected field $\delta_\parallel$ is not exactly the same as the radial correlation function $\xi(r,\mu=1)$ for finite angular pixel resolution. It is therefore important to study whether the changes in the correlation function due to the projection onto the radial direction alter the shape of the 2PCF on scales close to the BAO, since that could potentially affect the performance of our method. The correlation function of the projected field is

$$\begin{aligned}\langle\delta_\parallel(z_1)\delta_\parallel(z_2)\rangle &= \frac{1}{(\delta\Omega^2)^2}\int d\Omega_1^2\int d\Omega_2^2\,\langle\delta(t(z_1),r(z_1),\theta_1,\phi_1)\delta(t(z_2),r(z_2),\theta_2,\phi_2)\rangle\\ &= \frac{D(z_1)D(z_2)}{(\delta\Omega^2)^2}\int d\Omega_1^2\int d\Omega_2^2\,\langle\delta(t_0,r_1,\theta_1,\phi_1)\delta(t_0,r_2,\theta_2,\phi_2)\rangle\\ &= \frac{D(z_1)D(z_2)}{(\delta\Omega^2)^2}\int d\Omega_1^2\int d\Omega_2^2\,\xi\left(r_{12}(\Delta r,\bar r,\alpha)\right),\end{aligned} \tag{3.25}$$

where

$$\begin{aligned} r_{12} &\equiv |\Delta r|\left[1+2\left(\frac{\bar r^2}{\Delta r^2}-\frac{1}{4}\right)(1-\cos\alpha)\right]^{1/2},\\ \cos(\alpha) &= \sin\theta_1\sin\theta_2\cos(\phi_1-\phi_2)+\cos\theta_1\cos\theta_2,\\ \Delta r &\equiv r_1-r_2,\qquad \bar r=(r_1+r_2)/2.\end{aligned}$$

For finite $\delta\Omega^2$, however, the integral (3.25) is very complicated, and depends on the shape of the pixel used. In order to study projection effects in a qualitative way, we will make the following approximation:

$$\begin{aligned}\frac{\langle\delta_\parallel(z_1)\delta_\parallel(z_2)\rangle}{D(z_1)D(z_2)} &= \frac{1}{(\delta\Omega^2)^2}\int d\Omega_1^2\int d\Omega_2^2\,\xi\left(r_{12}(\Delta r,\bar r,\alpha)\right)\\ &\simeq \frac{1}{\alpha_r}\int_0^{\alpha_r} d\alpha\,\xi\left(r_{12}(\Delta r,\bar r,\alpha)\right),\end{aligned}$$

where α_r is a measure of the angular pixel size.

For a power-law correlation function $\xi(r)=\left[\frac{r_0}{r}\right]^\gamma$, and for small pixel resolutions ($1-\cos\alpha\sim\alpha^2/2$) this can be solved analytically as

$$\frac{\langle\delta(z_1)\delta(z_2)\rangle}{(D(z_1)D(z_2)} = g\left(\frac{\Delta r}{\bar r},\alpha_r\right)\xi(\Delta r), \tag{3.26}$$

where the correcting factor g is

$$g(x,\alpha_r) \equiv \frac{1}{\alpha_r}\int_0^{\alpha_r} d\alpha \left[1+\left(\frac{1}{x^2}-\frac{1}{4}\right)\alpha^2\right]^{-\gamma/2} = {}_2F_1\left[\frac{1}{2},\frac{\gamma}{2},\frac{3}{2},\left(\frac{1}{x^2}-\frac{1}{4}\right)\alpha_r^2\right].$$

For our purposes, it is interesting to study the small-scale and large-scale limits of this result:

- $g(x,\alpha_r\to 0)\to 1$: no effect for infinite resolution.
- $g(x\gg 1,\alpha_r)\to 1$: no effect on large scales.
- On small scales, the ratio of two correlations calculated with two resolutions is inversely proportional to the ratio of the resolutions:

$$\frac{g(x,\alpha_1)}{g(x,\alpha_2)} = \frac{\alpha_2}{\alpha_1} + \mathcal{O}(x).$$

- The tilt on small scales is modified from $-\gamma$ to $1-\gamma$:

$$\frac{d\,\ln g(x,\alpha)}{d\,\ln x} = 1 + \mathcal{O}(x).$$

Therefore, all projection effects seem to occur on small scales, and we can expect the large BAO scale to remain unchanged. The validity of this result has been more thoroughly studied and confirmed using the mock catalogue (see section 3.3.3).

3.3.2 Method to measure the radial BAO scale

As we did for the angular case, our method for the determination of the radial BAO scale relies on a parametrization of the radial correlation function. The suggested parametrization was inspired by the application of the Kaiser effect to the shape used for the angular correlation function. The full recipe to obtain the radial BAO scale as a function of redshift for a galaxy survey is as follows:

1. Divide the full galaxy sample in redshift bins.
2. Divide each redshift bin in angular pixels.
3. We consider pairs of galaxies inside the same pixel to be collinear, and therefore we compute the radial correlation function by stacking galaxy pairs in each angular pixel. We do not mix galaxies in different angular pixels.
4. Parametrize the correlation function using the expression:

$$\xi_{\parallel}(\Delta z) = A + Be^{-C\Delta z} - De^{-E\Delta z} + Fe^{-\frac{(\Delta z-\Delta z_{\rm BAO})^2}{2\sigma^2}} \tag{3.27}$$

   and perform a fit to $\xi_{\parallel}(\Delta z)$ with free parameters A, B, C, D, E, F, $\Delta z_{\rm BAO}$ and σ.
5. The radial BAO scale is given by the parameter $\Delta z_{\rm BAO}$. The BAO scale as a function of the redshift is the only parameter needed to apply the standard ruler method. The cosmological interpretation of the other parameters is limited, since this is an empirical description, valid only in the neighbourhood of the BAO peak.

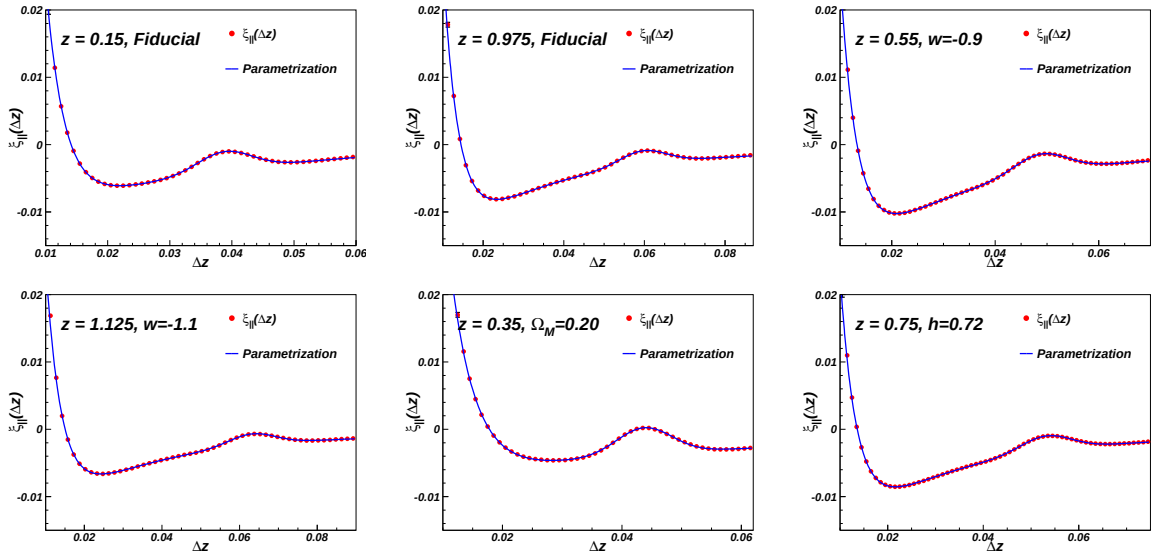

Figure 3.15: Some examples of the description of the different cosmological models by the proposed parametrization. Models do include redshift-space distortions and non-linearities using the RPT scheme. The parametrization is a very good description of all models at all redshifts, even for errors much smaller than the cosmic variance for the full sky. We have used this level of precision to ensure that the systematic errors associated with the theory (non-linearities, bias ...) are small.

6. As explained in the previous section, angular projection effects do not affect the BAO scale in the radial direction, unlike the angular case. Therefore $\Delta z_{\rm BAO}$ is directly used as an estimator for the radial BAO scale, without any correction for projection effects.

7. Fit cosmological parameters to the evolution of Δz_{BAO} with z.

In order to test the goodness of the parametrization (3.27), we have computed the radial correlation function for the 14 cosmological models described in table 3.1 in a redshift range from 0.2 to 1.5, always including redshift-space distortions and non-linearities as explained before. Then, we have applied the method to each model and each redshift. The parametrization describes the theory very accurately for all redshifts and cosmological models: we find values of the χ^2/ndof close to 1, and the probabilities of the fit lie between 0.98 and 1.00 for errors arbitrarily set to 2%. This error corresponds to a precision much better than the cosmic variance for the full sky. The parametrization is able to recover the correct radial BAO scale through the parameter $\Delta z_{\rm BAO}$ for the 14 cosmologies. We have used this level of precision because the systematic errors coming from theoretical effects (non-linearities, bias, fingers of god) only affect this method if they induce a disagreement between the models and the parametrization, giving then a wrong measurement of the BAO scale. If the description is good to this level of precision we guarantee a small contribution from these systematic effects, as is shown below. The calculation of the realistic errors is described in the next section. Some examples of the theoretical correlation functions can be seen in figure 3.15.

Once we have verified the validity of the parametrization on the theoretical predictions, we are ready to apply it in a more realistic case. For this purpose, in the next section we apply this algorithm to a mock galaxy catalogue obtained from a N-body cosmological simulation.

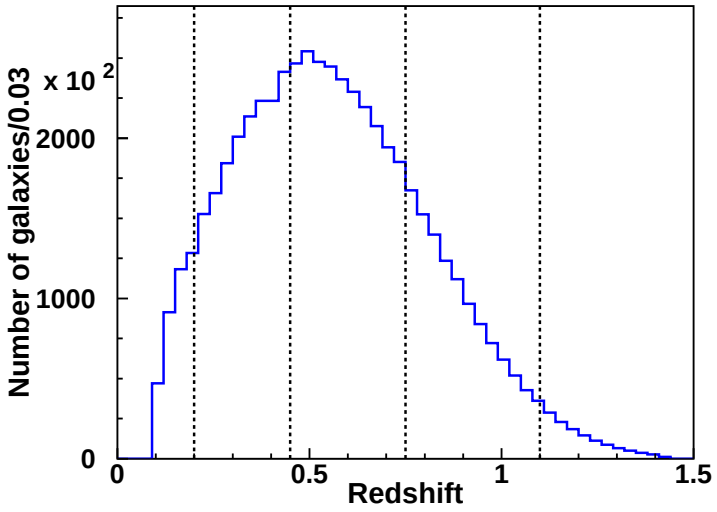

Figure 3.16: $N(z)$ in the MICE catalogue. The simulation contains 55 million galaxies in the redshift range $0.1 < z < 1.5$. The vertical dashed lines show the limits of the redshift bins used in the analysis.

3.3.3 Application to a simulated galaxy survey

We have tested the method to recover the radial BAO scale using a large N-body simulation capable of reproducing the geometry (e.g. area, density and depth) and general features of a large galaxy survey. The simulated data were kindly provided by the MICE project team, and consisted of a distribution of dark matter particles (galaxies, from now on) with the cosmological parameters fixed to the fiducial model of table 3.1. The redshift distribution of the galaxies is shown in figure 3.16. The simulation covers 1/8 of the full sky (around 5000 square degrees) in the redshift range $0.1 < z < 1.5$, and contains 55 million galaxies. This catalogue was extracted from an N-body simulation that is equivalent to the one used in the analysis of the angular BAO (section 3.2.4). However, this new dataset is more realistic, since it was produced from the lightcone output (and not at a comoving snapshot), and includes redshift-space distortions, which are fundamental in order to study the radial BAO feature. Data with similar characteristics will be obtained in future large spectroscopic surveys, such as DESpec [181], BigBOSS [182] or EUCLID [183].

It is important to note that the radial BAO determination needs a very large survey volume. We tried to extract the BAO peak from catalogues with smaller areas (200, 500 and 1000 deg^2), finding a very small significance (or no detection at all) in most cases. This is due to the fact that the statistical error related to the cosmic variance is specially large for the radial correlation function, and therefore it can only be reduced by increasing the volume explored. For this reason we have divided the simulation into 4 wide redshift bins, also shown in figure 3.16. We apply the method described in the previous section for each bin and obtain the correlation functions using the Landy-Szalay [176] estimator. The fits to each bin are shown in figure 3.17. The random catalogues must be generated taking into account the redshift distribution of the survey. Since shot noise is a relevant source of error for the radial correlation function (see figure 3.18), the random catalogues contained many times more particles than the data. The statistical significance of the BAO detection in the first bin is very low ($\sim 1.4\sigma$) and consequently was not considered to obtain cosmological constraints. Following the same approach of [1], we have computed the statistical significance of the detection by comparing the value of the amplitude of the Gaussian term (the F parameter)

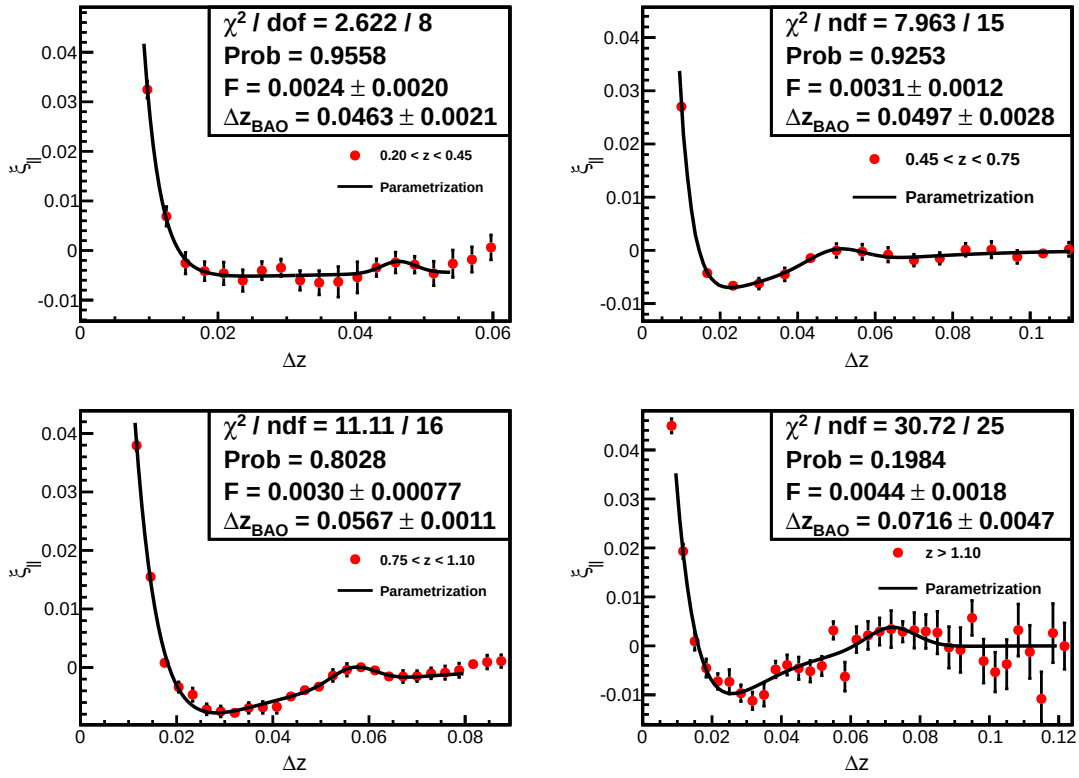

Figure 3.17: Radial correlation functions measured in the MICE simulation for the 4 redshift bins described in the text (dots) compared with the proposed parametrization (solid line). All the fits are good. The statistical significance of the BAO detection in the first bin is very low, and it is not used to set cosmological constraints.

with its error.

The statistical errors were computed theoretically using a Gaussian covariance matrix (see appendix B.2.1 for further details - eq. (B.16)). This estimation depends on the model used for the simulation, but we expect a small variation of the error with the cosmological model. In any case, to keep the method fully model-independent we have validated the calculation of the Gaussian covariance matrix comparing it with the estimation obtained by dividing the catalogue into several equivalent regions. The correlation function is measured in each of these regions and the statistical error is determined as the r.m.s. of these measurements. The errors for the survey are then estimated by scaling this result with the full area of the survey. Both estimations agree in the region of interest for the BAO scale measurement, as is shown in figure 3.18. The disagreement for small scales may come from an incomplete description of non-linear shot noise and clustering. On large scales the limit of the redshift bin is reached, and boundary effects are not taken into account in the theoretical calculation.

It is important to remark that the volume which the radial BAO determination needs is very large. We were not able to obtain a significant observation for the first redshift bin and the contribution to the total error of the cosmic variance and the shot noise is comparable for all the other bins (as is also shown in figure 3.18). This is not surprising, since the number of galaxy pairs that are considered collinear is very small.

The measured values of the BAO scale as a function of redshift can be seen in figure 3.19 as dots. The results for 3 alternative bins shifted with respect to the nominal ones are also shown as triangles. These were not used to obtain cosmological constraints, but are shown

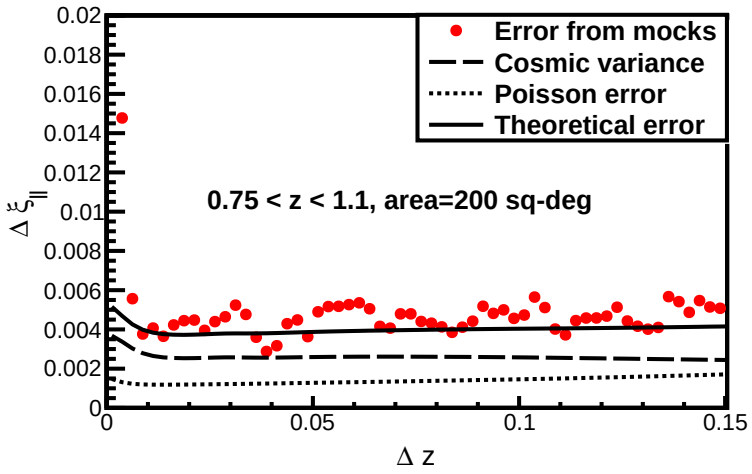

Figure 3.18: Comparison of the different estimates of the error in the radial correlation function. The estimates are MC samples (dots) and theoretical calculation (solid line). They agree in the region of interest for the BAO analysis. The disagreement at low scales comes from the incomplete description of the non-linearities, where the mode-mode coupling effect is neglected, from boundary effects and from unmodelled non-linear shot noise, but does not affect the BAO scale measurement since it is outside the fitting region. The different contributions to the total error are shown. They come from Poisson shot noise (dotted line) and cosmic variance (dashed line). The contribution of the Poisson shot noise is not negligible.

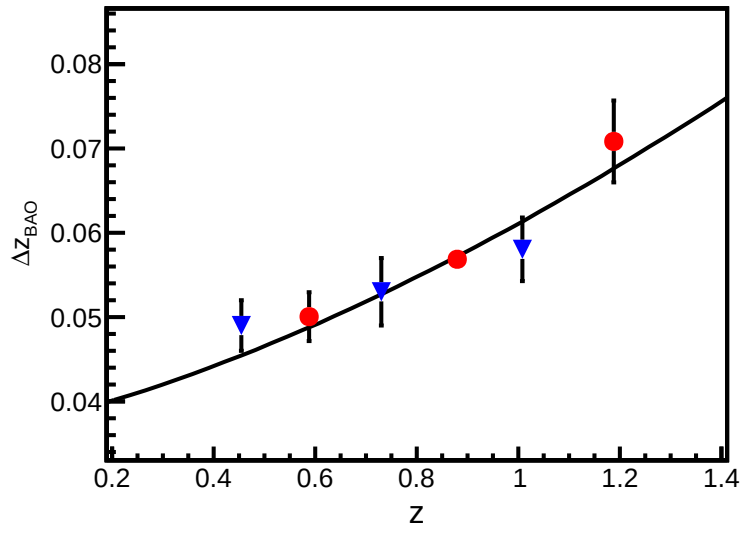

Figure 3.19: Measured radial BAO scale as a function of the redshift in the MICE simulation using the proposed method. Dots are the nominal bins and triangles correspond to displaced bins, and are measured only as a cross-check. All measurements are in good agreement with the true cosmology (solid line).

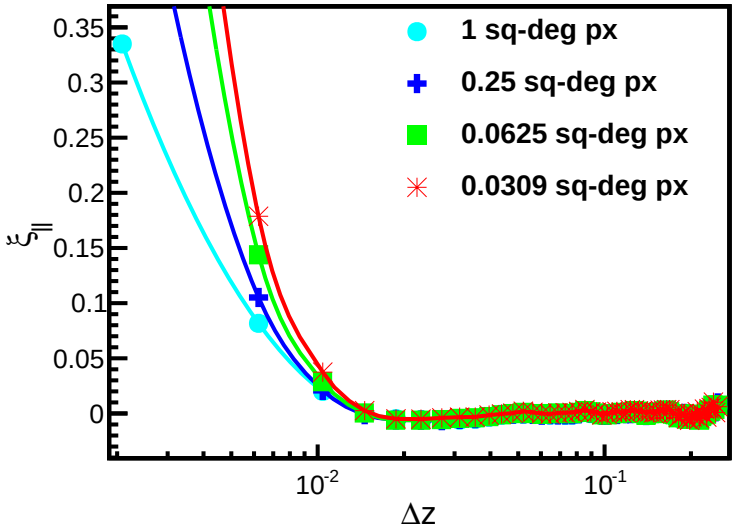

Figure 3.20: Effect of the angular pixel size on the radial correlation function. The finite size of the angular pixel induces a change in the slope of the correlation function at small scales. The effect is shown for pixels of sizes 0.0309 (stars), 0.0625 (squares), 0.25 (crosses) and 1 (dots) degrees. The change in the slope arises from the smoothing effect produced by the inclusion in the calculation of galaxy pairs which are not exactly collinear. This effect does not affect the determination of the BAO scale.

for illustrative purposes and to verify the correct behaviour of the method. For each bin we have used the mean redshift weighed by the galaxy number density to obtain the prediction of the model.

Systematic errors

We have studied the main systematic errors that affect the determination of the radial BAO scale using this method. A specific systematic is associated with the angular projection of the galaxy number density. Other sources are generic and will be present in any BAO study, namely, the influence of the non-linearities, the starting and end point of the fit and the possible influence of the galaxy bias in the measurement.

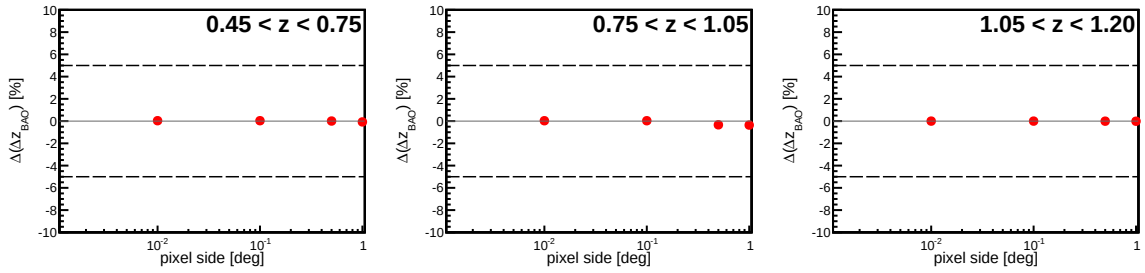

Figure 3.21: Variation of the radial BAO scale determination as a function of the angular pixel size for different redshifts. Results are stable, and the maximum variation is always of a few parts per mille, very well below 1%, for a range of pixel sizes covering two orders of magnitude. Dashed lines indicate the size of the statistical error for the mock catalogue, including Poisson shot noise and cosmic variance.

- **Angular projection.** As explained in section 3.3.1, by allowing a finite angular separation between collinear galaxies, we are effectively projecting the galaxy number density in the radial direction in a finite angular bin. This is in direct analogy with the projection in the redshift bin of the angular correlation function. However, unlike in this case, for small enough angular pixels the effect of this projection is limited to small values of Δz, far away from the BAO scale. This is explicitly shown in figure 3.20, where we have calculated the radial correlation function in angular pixels with different resolutions. Larger angular pixels produce a smoothing of the radial clustering pattern, however the scale of this smoothing is limited by the angular resolution, and for small enough pixels the BAO scale remains unaffected.

  In order to quantify the influence of this effect on the determination of the BAO scale as a systematic error, we have repeated the full analysis for different pixel sizes. The results are shown in figure 3.21. The radial BAO scale is recovered with high precision for any angular pixel size, even for sizes as large as 1 degree, which corresponds to a range of two orders of magnitude. The associated systematic error can be estimated to be $\delta(\Delta z_{BAO}) = 0.20\%$.

- **Non-linearities.** The error due to the uncertainty in the goodness of the description provided by the parametrization for different theoretical effects (non-linearities at the scale of the BAO peak) has been estimated obtaining a global error of 0.10%. This was calculated in a conservative way as the difference between the Δz_{BAO} measured using linear and non-linear $\xi_{\parallel}(\Delta z)$, for the same redshift bins of the analysis. Non-linearities are computed using the RPT formalism excluding mode-mode coupling, which only affects small scales, far enough from the BAO scale. The contribution of this source to the total systematic error can be estimated as $\delta(\Delta z_{BAO}) = 0.10\%$.

- **Galaxy bias.** At the BAO scales it is a good approximation to consider that the bias is scale-independent [178, 163]. It affects the radial correlation function not only as an overall normalization, but also through its effect in the redshift-space distortions. Bias can influence the determination of the BAO scale only through the changes in the goodness of the parametrization of the correlation function for different biases. In order to estimate the contribution of the galaxy bias to the measurement of the radial BAO scale, we have repeated the analysis with different values of the bias, to obtain the propagation of the uncertainty in the galaxy bias for the selected galaxy population to the measured value of the radial BAO scale. The effect of bias on the shape of the radial correlation function can be absorbed by our parametrization, and the influence on the peak position is small. The associated systematic error is estimated to be $\delta(\Delta z_{BAO}) = 0.15\%$.

  Moreover, we have tested the effect of a scale-dependent bias, introducing artificially the effect in the correlation functions, using an approximated Q-model with the determination of parameters of [178]. The variation of b with Δz in the fitted region of $\xi_{\parallel}(\Delta z)$ ranges from 1% to 6%, but the measurement of the BAO scale is insensitive to these changes. We estimate the systematic error in the presence of a scale-dependent bias as $\delta(\Delta z_{BAO}) = 0.20\%$.

- **Limits of the fit.** As was done for the angular BAO scale, in order to compute the systematic error associated with the parametrization, we have done some further analysis on the theoretical radial correlation functions in the same bins as those used in the study of the MICE catalogue. The error associated with the method comes from the

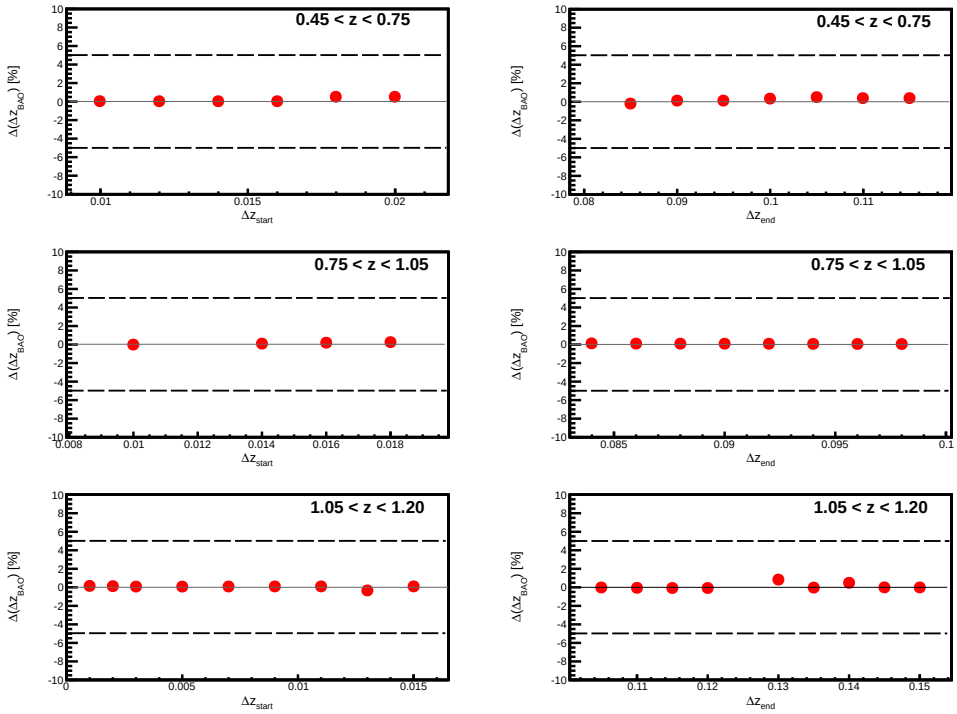

Figure 3.22: Evolution of the error on Δz_{BAO} as a function of the starting and end point of the fitted region. Results are stable, confirming that the systematic error is small. Dashed lines indicate the size of the statistical error, including Poisson shot noise and cosmic variance in 1/8 of the sky.

possible influence over the value of Δz_{BAO} of the range of Δz used to perform the fit. To evaluate the error, we have varied this range for the 3 redshift bins where we have a significant detection of the BAO scale. In the decision of the range to be fitted, we have to choose a starting point at scales smaller than the BAO peak and an end point after the peak. The results for different fitting ranges can be seen in fig. 3.22, where the value of Δz_{BAO} is shown for different starting points and end points of the fit, for the 3 redshift bins. In all cases, the uncertainty is of the order of 0.1%, which we use as the associated systematic error.

The different sources of the systematic errors are completely independent, and therefore we can compute the total systematic error by adding these contributions in quadrature, leading to a value of $\delta_{SYS}(\Delta z_{BAO}) = 0.33\%$. There exist some other potential systematic errors: the gravitational lensing magnification, which introduces a small correlation between redshift bins, or those mainly associated with the instrumental effects which could affect the galaxy sample. However, these effects are expected to be very small and we have neglected them in this analysis.

Cosmological constraints

The evolution of the measured radial BAO scale, including the systematic errors, with redshift is shown in figure 3.19. The cosmological model of the simulation is the solid line. The recovered BAO scale is perfectly compatible with the true model, demonstrating that the method works.

The corresponding constraints on the cosmological parameters are depicted in figure 3.23,

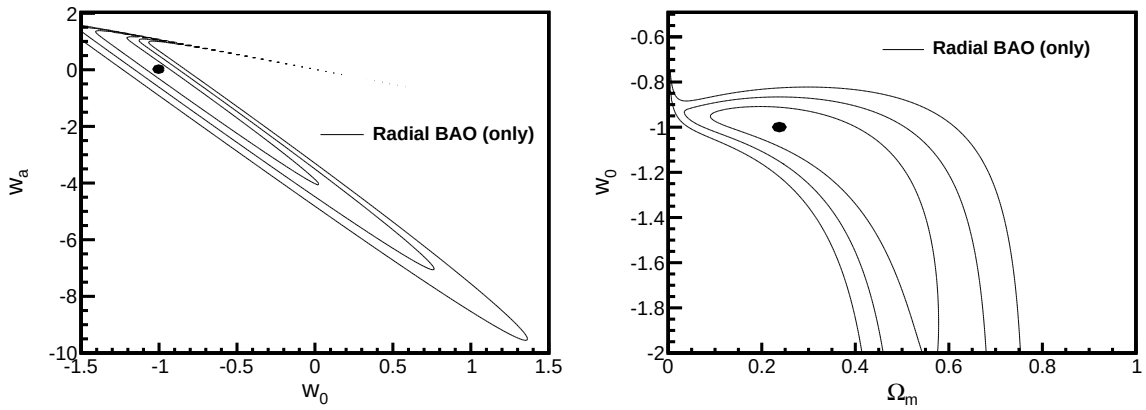

Figure 3.23: Contours at 1σ, 2σ and 3σ C. L. on (w_0, w_a) (left) and on $(\Omega_{\rm M}, w_0)$ (right) obtained from the analysis of the radial BAO scale. The dot shows the value of the parameters for the MICE cosmology. No combination with any other cosmological probe is included. The other parameters have been fixed to the values of the MICE cosmology.

where the contours for 1σ, 2σ and 3σ C. L. in the (w_0, w_a) plane are shown in the left panel and the right panel shows the same contours in the $(\Omega_{\rm M}, w_0)$ plane. All other parameters were kept fixed to their fiducial values. The MICE cosmology is recovered, and the plot shows the sensitivity of the radial BAO scale alone, since no other cosmological probe is included in these constraints.

3.3.4 Discussion

The method outlined in the previous section to measure the angular BAO scale can be extended to detect the BAO feature in the distribution of galaxies along the line of sight. As in the angular case, this methodology is adapted to the observational characteristics of galaxy surveys, where only angular coordinates and redshift are measured for each galaxy. The sound horizon scale in the radial direction can be recovered from the non-linear radial correlation function, however a very large survey volume is necessary to obtain a significant detection, due to the large effect of cosmic variance on this measurement. Therefore, detecting the radial BAO scale is only limited by the survey volume and number density, since, as we have discussed, the systematic uncertainties associated with this measurement are very small. On the other hand, the method is fully cosmology-independent, since it relies only on observable quantities and, consequently, its results can be used to constrain any cosmological model.

The method has been tested with a mock catalogue built upon a large N-body simulation provided by the MICE collaboration. The true cosmology is recovered within 1-σ from the measurements of $\Delta z_{\rm BAO}$. The main sources of systematic error have been included in this study, and we find that the method is very promising and robust against these. Although the analysis was done on dark matter particles instead of galaxies, we have argued that the effect of galaxy bias does not limit the sensitivity of the method.

These results were obtained assuming that the the galaxy redshifts could be measured with very good accuracy, and therefore only hold for spectroscopic surveys. It would be interesting to study the dependence of the significance of this measurement with the accuracy in the determination of the redshifts, in order evaluate whether narrow-band photometric redshifts (e.g. PAU [146], $\sigma_z \sim 0.003$) could be used to measure the radial BAO scale. However, we have left this study for the future.

3.4 Combining angular and radial BAO information

As we have argued, two different complementary BAO observables can be used: in the plane transverse to the line of sight, the angular BAO scale gives information about the angular diameter distance as a function of redshift, and along the radial direction the BAO can be used to trace the evolution of the expansion rate

$$\theta_{\rm BAO}(z) = \frac{r_s}{(1+z)\, d_A(z)}, \quad \Delta z_{\rm BAO} = r_s\, H(z) \tag{3.28}$$

In a galaxy redshift survey, what is observationally accessible is the distribution of galaxies in the space of angles and redshift, and not their Cartesian coordinates. In order to measure the 3D correlation function $\xi(\mathbf{r})$ from a galaxy survey, redshifts must be translated into comoving distances using some fiducial cosmological model, and by manipulating the data in this manner it is not clear that the measurement of the BAO will be fully model-independent. In fact, the purely observable coordinates (angles and redshifts) are directly related to the two BAO scales $(\theta_{\rm BAO}, \Delta z_{\rm BAO})$, and therefore it is natural to try to measure them from the angular and radial correlation functions separately. A fully model-independent approach to LSS, making use only of pure observables has been recently advocated by different groups [184, 185, 186].

In the two previous sections we have described two methods to measure the radial and angular correlation functions, and to extract $\theta_{\rm BAO}$ and $\Delta z_{\rm BAO}$ from them in a model-independent way. In this section we will evaluate the constraining power of combining both methods on the same dataset (section 3.4.1). Without a closer study it is not clear whether the standard analysis, which makes use of all the available galaxy pairs, is more powerful than our combined approach. Our results combining angular and radial measurements were compared with those obtained with the standard approach in section (see section 3.4.2). The main results from this comparison are summarized in section 3.4.3.

3.4.1 Combined angular and radial BAO measurements

We used the method described in section 3.2 to measure the angular BAO scale from the mock catalogue used in the analysis of the radial BAO described in the previous section. For this, redshift bins of width $\Delta z = 0.1$ were used from $z = 0.3$ to $z = 1.3$, making a total of 10 measurements of $\theta_{\rm BAO}$. It is important to note that we have assumed spectroscopic redshifts, so the effective bin width is not increased by the photometric redshifts, no systematic photo-z errors must be used and no correlations between bins are necessary. This improves the constraints on the cosmological parameters from the results quoted in section 3.2, where photometric redshifts were used.

We combined these 10 angular measurements with the 3 values found in the previous section for $\Delta z_{\rm BAO}$ in order to obtain tighter cosmological constraints. The separate and combined contours in the $(\Omega_{\rm M}, w_0)$ and (w_0, w_a) planes are shown in figure 3.24. As was done in the previous two sections, all other cosmological parameters were fixed to their fiducial values. As expected, the joint constraints are significantly better, and especially the right panel of figure 3.24 shows the complementarity of both observables.

As is usually done with BAO measurements, we have combined them with CMB data using the WMAP7 covariance matrix [32] and assuming the measurements correspond to the MICE cosmology. This combination was done following the procedure described in [187]. The corresponding contours for the same sets of parameters as before are shown in figure 3.25. As expected, the constraints improve significantly when including CMB data.

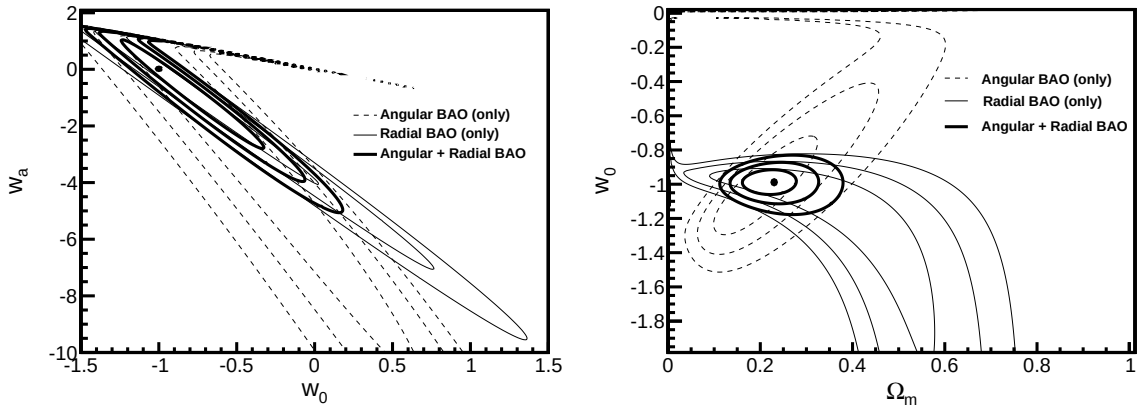

Figure 3.24: Contours at 1σ, 2σ and 3σ C. L. on the plane (w_0, w_a) (left) and on the plane $(\Omega_{\rm M}, w_0)$ (right) from radial BAO (thin solid lines), angular BAO (thin dashed lines) and the combination of both (thick solid lines). The dot shows the value of the parameters for the MICE cosmology. No other cosmological probe is included in this result, showing the high sensitivity that the BAO standard ruler can achieve. The other parameters have been fixed to their values in the MICE cosmology.

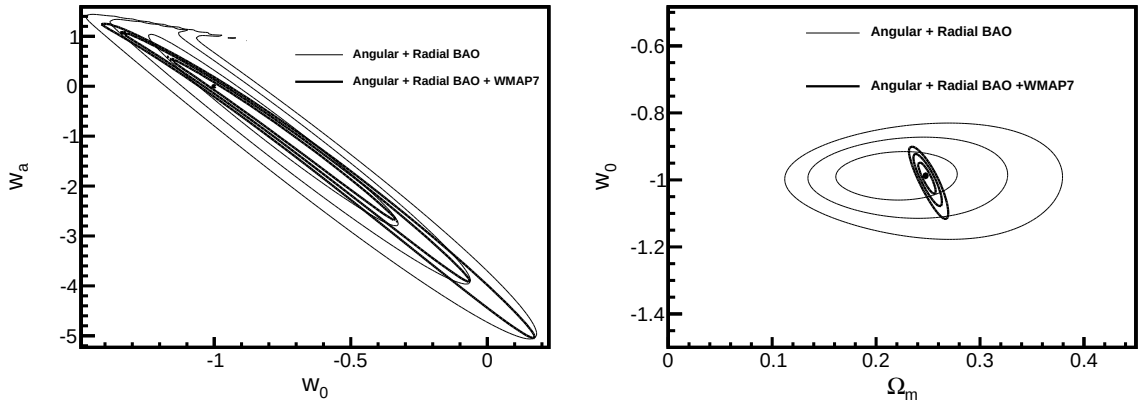

Figure 3.25: Contours at 1σ, 2σ and 3σ C. L. on the plane (w_0, w_a) (left) and on the plane $(\Omega_{\rm M}, w_0)$ (right) from the combination of radial BAO information (thin solid lines), and adding also the CMB measurements (thick lines). The dot shows the value of the parameters for the MICE cosmology. The covariance matrix of WMAP7 has been used, while the central value of the measurement has been taken at the MICE cosmology. The other parameters have been fixed to their values in the MICE cosmology.

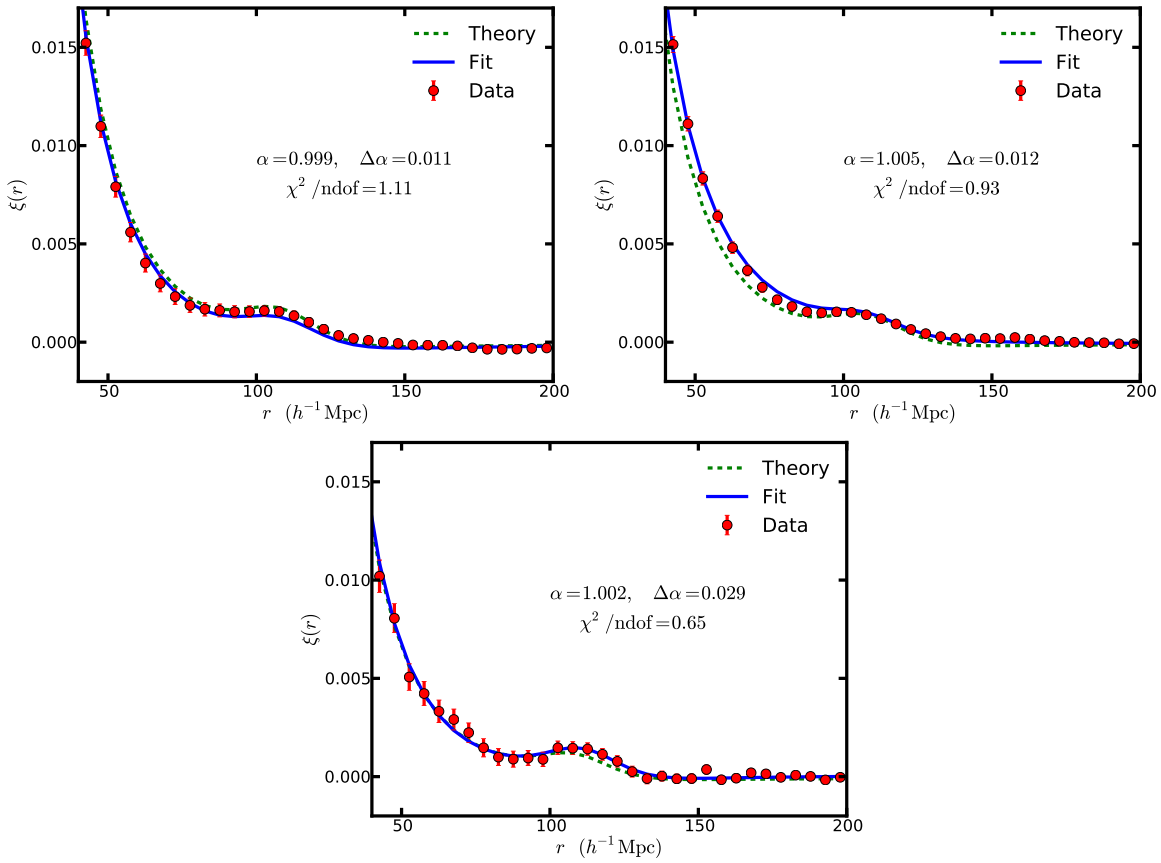

Figure 3.26: Monopole correlation functions measured from the MICE mock catalogue. The solid blue lines correspond to the best fit using the model in eq. (3.1). Dashed green lines correspond to the theoretical prediction for the MICE cosmology, used as a template.

3.4.2 Comparison with the standard analysis

In order to compare our results, obtained combining radial and angular BAO, with the standard approach of measuring the position of the sound horizon scale in the monopole 2PCF, we performed on our mock catalogue the same analysis that was carried out to obtain the latest BAO detection by the BOSS collaboration [30] (BOSS-BAO from here on). For this we calculated the three-dimensional correlation function $\xi(r)$ using the Landy and Szalay estimator in the three wide redshift bins used for the analysis of the radial BAO ($0.45 < z < 0.75$, $0.75 < z < 1.10$ and $1.10 < z$). We chose these wide bins in order to maximize the number of pairs that contribute to the measurement of the monopole, since this is one of the key advantages of the standard method. In order to calculate $\xi(r)$, redshifts must be translated into distances. Even though it would be extremely interesting to study the dependence of the final result on the choice of the fiducial cosmology to perform this task, we have not done so, and we have used the true cosmology of the MICE simulation to ensure that our results will not be biased by this choice.

The covariance matrix for the monopole was calculated using the Gaussian approach described in section B.3.1 (eq (B.28)). This calculation was, again, validated by comparing it with the errors computed from subsamples of the total catalogue, and both estimations were found to be compatible within the range of scales needed for the analysis.

Once the 2PCFs and the covariance matrices are calculated, we fit the model given by

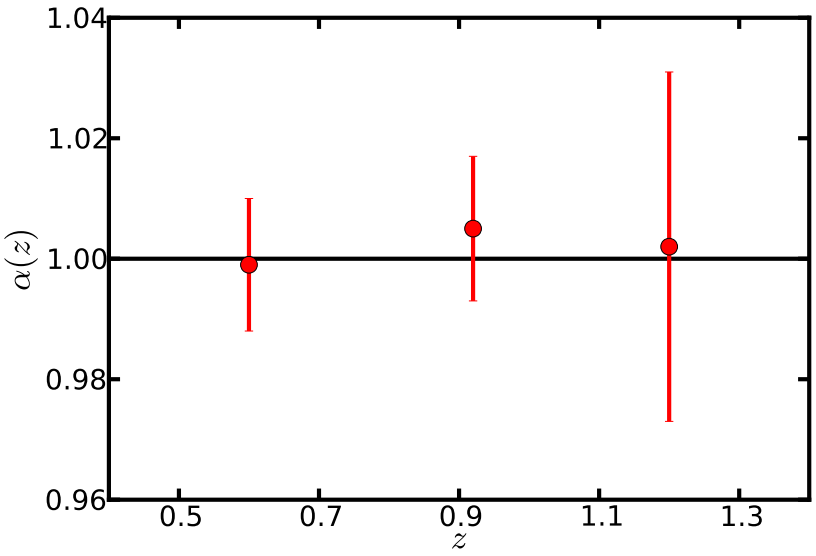

Figure 3.27: Values of the scaling parameter α measured from the three bins of the MICE mock catalogue. The input cosmology ($\alpha \equiv 1$) is recovered well within errors.

equations (3.1) and (3.2) to the data, where the template for the theoretical correlation function was calculated from the `CAMB` linear power spectrum for the MICE cosmology, and corrected for non-linearities via the RPT damping factor. We are mainly interested in the fitting parameter α, which relates real and fiducial scales (see eq. (3.6)). Since the true cosmology was used to translate redshifts into distances, the value of α must be compatible with 1 (within errors). The statistical uncertainty in α was calculated following the same method used in BOSS-BAO: for each fixed value of α, we find the best fit parameters (B, a_i), and calculate the $\chi^2(\alpha)$ for those. The probability of a given value of α is $p(\alpha) \propto \exp(-\chi^2(\alpha)/2)$, and the uncertainty in α is found using this distribution as $\sigma_\alpha^2 \equiv \langle\alpha^2\rangle - \langle\alpha\rangle^2$.

Figure 3.26 shows the correlation functions measured in the three redshift bins (red dots) together with the best fit to the model in eq. (3.1) (blue solid lines) and with the theoretical correlation function used as a template for the fit (green dashed lines). The error bars correspond to the statistical errors. In all cases the values of the χ^2 for the best fit parameters are close to 1. The values of α extracted from the fit, together with their errors are shown in figure 3.27. As can be seen, the true cosmology ($\alpha = 1$) lies well within the statistical uncertainties.

We have not studied the different sources of systematic errors for this measurement, and no systematic contribution has been added to the errors. On the one hand this provides a more conservative comparison with our approach, since the results quoted in the previous section do contain systematics. On the other hand, there exist several potential systematics that are specific for the standard method, such as the effect of the fiducial cosmology used to obtain the three-dimensional positions of the galaxies, or the choice of template used to perform the fit. Studying this effect would be extremely interesting, but we have postponed this analysis for a future work. As we have seen before, the systematic errors that are common to both approaches (bias, RSDs, non-linearities, fitting limits) are clearly subdominant compared to the statistical uncertainties [188].

No reconstruction techniques were used in this analysis [160], and the widening of the BAO feature due to non-linearities was included in the fitting template. The reason for this is twofold: first, the results with and without reconstruction quoted by BOSS-BAO show a very

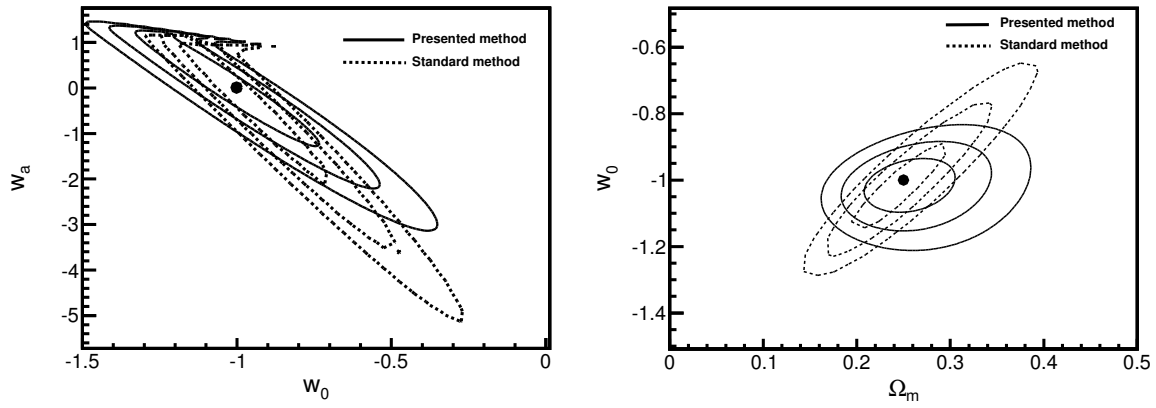

Figure 3.28: Contours at 1-σ, 2-σ and 3-σ confidence level in the planes (w_0, w_a) (left panel) and $(\Omega_{\rm M}, w_0)$ (right panel). The solid contours correspond to our combination of independent radial and angular BAO information, while the dashed lines correspond to the results drawn from the standard analysis of the angle-averaged BAO.

small improvement when using this technique. Secondly, implementing BAO reconstruction requires manipulating the positions of the galaxies. As we have argued, one of the caveats of the standard method is the fact that it requires transforming the data according to some fiducial cosmological model, which might bias the final result. Reconstruction techniques would alter the data further in a cosmology-dependent way, which we prefer to avoid.

The cosmological constraints drawn from the values of α measured from the correlation functions are shown in figure 3.28 in the $(\Omega_{\rm M}, w_0)$ and (w_0, w_a) planes. The figure also shows the contours corresponding to the combination of radial and angular information, described in the previous section. Both plots show that the constraining power of both methods is very similar. There is a degenerate direction in the $(\Omega_{\rm M}, w_0)$ plane for the standard method, which coincides with the orientation of the contours for the angular BAO shown in figure 3.24 (right panel). This is a reasonable result: most of the information in the angle-averaged BAO signature comes from the angular part, since there are two transverse dimensions and only 1 longitudinal. On the other hand, our combined approach seems to be able to obtain better constraints on the evolution of the dark energy equation of state. This could be due to the fact that the radial BAO enables us to measure the evolution of the expansion rate alone, which is a local quantity, unlike the angular diameter distance, which is an integrated one depending on the expansion history to $z = 0$.

3.4.3 Discussion

As we have shown, important constraints on the cosmological parameters can be obtained by combining radial and angular BAO data alone. We have compared the constraining power of our method (i.e. measuring the angular and radial BAO scales separately using only pure observables) with the standard approach, which measures the BAO scale in the 3D angle-averaged correlation function. As has been shown, both methods seem to have a similar constraining power, but it is worth pointing out the differences between them in the light of the present study.

The standard approach has many advantages. First, it makes a very efficient use of the galaxy sample, since no galaxy pair is thrown away. Also, by recovering the three-dimensional positions of the galaxies (albeit through a fiducial cosmological model), it is possible to implement reconstruction techniques which may improve the sensitivity of the

measurement. Finally, since the method has been widely used and tested in many practical cases, it has been thoroughly explored, and the large amount of literature about it makes it easier to implement. However, this method of analysis has some caveats which cannot be ignored. Since this approach requires the use of a fiducial cosmology on the data, it is not clear that the results will not be biased towards this model in some cases. Also, the method used to extract the BAO scale from the correlation function does not recover it from an independent contribution to the fitting model (such as the Gaussian term in our approach). Instead, it involves fitting the whole shape of a template for the theoretical correlation function. While this may improve the accuracy of the measurement, it also makes it more prone to systematic uncertainties if the template used is not correct (for example, if the fiducial cosmology used for the template is wrong). Most importantly, this method is able to extract information about the BAO from a combination of the radial and angular contributions, and does not measure each of them independently. As we have shown, this fact can reduce the sensitivity to certain parameters, such as the evolution of the equation of state of dark energy. An independent measurement of the radial and angular BAO scales has been recently claimed by the BOSS collaboration [189], using a method similar to the standard one we have described (i.e.: fiducial model and template fitting). Surprisingly, the sensitivity to the cosmological parameters does not increase significantly with respect to the volume-averaged measurement.

The most important aim of our work in the papers [1, 2] was to develop a method to recover the radial and angular BAO scales independently through the use of pure observables, and therefore in a fully model-independent way. In these last 3 sections we have described the method and shown that it can be used successfully in LSS surveys. The results obtained with this method can be directly used to constrain any cosmological model, (as long as we are able to accurately link theory and observations). Furthermore, the sensitivity of the method is comparable to that of the standard approach, even outperforming it in some regions in parameter space (see figure 3.28). Our method is not free of disadvantages. We discard some galaxy pairs (those which are in between radial and angular), that may contain useful information. Also, since it has not been widely used, it is not thoroughly covered in the literature, and it may be harder to implement. Finally, since the three-dimensional galaxy positions are not accessible, it is not clear whether BAO reconstruction techniques could be implemented in this analysis in order to improve its sensitivity.

CHAPTER 4

Structure formation in LTB models

The standard cosmological model, the commonly accepted picture describing the origin and evolution of the Universe on large scales, is based on a minimal set of assumptions

- The Cosmological Principle. This is a basic philosophical premise that describes the Universe as a spatially maximally symmetric system, where all points in space are statistically equivalent.
- General relativity describes the evolution of this homogeneous patch in terms of the FRW metric, which predicts that the Universe must be expanding.
- The inflationary paradigm explains the production of large flat regions of space at early times, as well as the initial conditions of the perturbations around these homogeneous configurations, which arise from quantum fluctuations on the density field.
- Gravitational collapse explains the growth of these fluctuations, which become the highly non-linear structures that we observe at late times. The spectrum of these perturbations can only be understood if most of the matter content of the Universe is in the form of a non-interacting dust component (cold dark matter).
- The apparent accelerated expansion of the Universe, deduced from the different measurements of the redshift-distance relationship, implies the existence of an unknown energy component with negative pressure (dark energy). The simplest model for this component is vacuum energy density (cosmological constant), which is allowed and expected in quantum field theory (although, as we have seen, its expected value differs greatly from the measured one).

The different issues concerning the standard description of dark energy as the source of cosmic acceleration have been thoroughly discussed in the literature (the fine-tuning problem, the coincidence problem...) and have driven the construction of many alternative DE models by those not contented with anthropic arguments. Nevertheless, the observations of the distance-redshift relation can only be interpreted, within the framework of homogeneous models, as evidence for an accelerated expansion.

However, it must be noted that the validity of the CP has only been directly verified within a range of scales. The galaxy distribution at late times seems to reach the homogeneous regime at scales of $\mathcal{O}(100)\,\mathrm{Mpc}/h$, however very few observations exist on larger scales, except for those related to the CMB. Indeed the CMB is extremely isotropic, showing temperature

fluctuations with amplitudes smaller than 1 part in 10^5, however statistical isotropy around a given point and the maximal symmetry postulated by the CP (homogeneity + isotropy) are different mathematical properties, and the latter is only granted if the former is realized around every point in space.

These issues have motivated the appearance of a new set of models which aim to probe the validity of the CP as well as to test whether, by dropping this premise, one is able to describe the observed distance-redshift data without DE. In these models one explores the possibility that we might be living close to the centre of a very large ($\mathcal{O}(1)\,\mathrm{Gpc}/h$) underdense region of space (a void). For these models to be compatible with the observed dipole of the CMB and the isotropy of the galaxy distribution, this region must be spherically symmetric and we must be very close to its centre. Hence the only symmetry that remains in these models is isotropy around the central observer, while all observables may vary with the distance to it. Such systems are described in GR through the Lemaître-Tolman-Bondi (LTB) metric, which these models are named after.

It is easy to understand why these models can potentially explain the observed redshift-distance relation. The key is to realise that most astronomical observations are carried out in the lightcone. Thus, while a variation with redshift can only be interpreted as a variation with time in terms of a homogeneous model, for an LTB model it is a combination of temporal and radial variations:

$$\frac{d}{dz} = \frac{dt}{dz}\,\partial_t + \frac{dr}{dz}\,\partial_r.$$

Hence, an observation of an expansion rate $H(z) < H_0$ at $z > 0$ is interpreted as a positive acceleration for a homogeneous model, while this is not necessarily so in an LTB model in which the expansion rate decreases with r (as happens in void models).

In the rest of this chapter we will present the key mathematical results (section 4.1) and current observational constraints (section 4.2) regarding LTB models. Then we will describe our contribution to their study: the development of the first (and so far only) N-body simulations of LTB void models (section 4.3), which have been applied to the analysis of halo abundances and orientations (section 4.4). Further details about the mathematical description of LTB models can be found in appendix C.

4.1 Lemaître-Tolman-Bondi models

4.1.1 The LTB metric

LTB models describe, within the framework of GR, spatially inhomogeneous universes that are isotropic with respect to a given point. It can be shown [12] that a 4-metric with maximally symmetric spatial 2-subspaces ($\mathcal{S}^2$) in comoving coordinates must take the form

$$ds^2 = -dt^2 + X^2(t,r)\,dr^2 + A^2(t,r)\,(d\theta^2 + \sin^2\theta\,d\varphi^2), \tag{4.1}$$

For most practical purposes we will see in sec. 4.1.4 that one can write

$$X(t,r) = \frac{A'(t,r)}{\sqrt{1-k(r)}}, \tag{4.2}$$

where ($'$) denotes ∂_r. Thus

$$ds^2 = -dt^2 + \frac{(A'(t,r))^2}{1-k(r)}\,dr^2 + A^2(t,r)\,d\Omega^2. \tag{4.3}$$

We see that we recover the FRW solution in the case

$$A(t,r) = r\,a(t), \quad k(r) = k\,r^2. \tag{4.4}$$

Just like gauge freedom allows us to impose an arbitrary initial condition $a(t_0) = a_0$ in the FRW case, here we can set

$$A(t_0,r) = A_0(r). \tag{4.5}$$

Generally one takes $A_0(r) = r$, which is analogous to $a_0 = 1$.

By inspection of the metric (4.3), it is easy to realise that LTB models have two different expansion rates in the radial (L) and transverse (T) directions

$$H_L \equiv \frac{\dot{A}'}{A'} \quad H_T \equiv \frac{\dot{A}}{A}. \tag{4.6}$$

This is a consequence of the lack of homogeneity: isotropic expansion can only occur at $r = 0$ or $r \to \infty$.

4.1.2 Redshift

Assuming that the observer is located at (or very close to) the centre of the void, we are interested in calculating the trajectories of radially infalling photons. From the metric, setting $d\Omega = 0$ we obtain the constraint

$$\dot{t}^2 - \frac{A'^2}{1-k(r)}\,\dot{r}^2 = \dot{t}^2 - A'^2\dot{\chi}^2 = \epsilon, \tag{4.7}$$

where $\epsilon = 0$ for $m = 0$ and $\epsilon = 1$ for massive particles (this can also be derived from the general geodesic equation. See section C.1). Here the dot ($\dot{}$) indicates differentiation with respect to an affine parameter.

From this result, the trajectory can be parametrized in terms of the redshift suffered by the photon [63], yielding the redshift equations

$$\frac{dt}{d\log(1+z)} = -\frac{A'(t(z),r(z))}{\dot{A}'(t(z),r(z))}, \quad \frac{dr}{d\log(1+z)} = \frac{\sqrt{1-k(r(z))}}{\dot{A}'(t(z),r(z))}. \tag{4.8}$$

These can be integrated to give the functions $r(z)$ and $t(z)$ with the initial conditions $r(z = 0) = 0$, $t(z = 0) = t_0$.

4.1.3 Distances

Using the same arguments and definitions traditionally used in the FRW case we can calculate the different distances to an observer in a LTB universe. It is rather straightforward to find closed expressions for them when the observer is located at the centre of symmetry, but a more careful treatment is needed for off-centred observers.

- **Proper (physical) distance.** This is defined as the distance measured by adding up the relative infinitesimal distances simultaneously measured by a chain of comoving observers placed in line between us and the point of interest (i.e.: the equivalent of $a\,\chi(a)$ in FRW):

$$d_p(r,t) = \int_0^r \frac{A'(r',t)}{\sqrt{1-k(r')}}\,dr'. \tag{4.9}$$

- **Angular diameter distance.** For an object of proper length l subtending an angle $\Delta\theta$ we define the *angular diameter distance* by

$$d_A = \frac{l}{\Delta\theta}. \tag{4.10}$$

Since $l = A(t_E, r_E)\,\Delta\theta$,

$$d_A(z) = A(t(z), r(z)) \tag{4.11}$$

- **Luminosity distance.** Since LTB models are based on a metric theory, the luminosity distance is related to the angular diameter distance by $d_L(z) = d_A(z)(1+z)^2$, hence

$$d_L(z) = (1+z)^2\, A(t(z), r(z)) \tag{4.12}$$

4.1.4 Einstein's equations

The energy-momentum tensor

Even though our aim is to use LTB models as an alternative explanation to the dimming of supernovae without using any kind of dark energy, and therefore in our case the energy momentum tensor is that of a pure CDM fluid in comoving coordinates:

$$T^{\mu}{}_{\nu} = -\rho_{\rm M}(r,t)\,\delta^{\mu}_{0}\delta^{0}_{\nu}, \tag{4.13}$$

we would like to take a more general approach here and include a DE component in the form of a cosmological constant, whose density must be, therefore, homogeneous:

$$T^{\mu}{}_{\nu} = -\rho_{\rm M}(r,t)\,\delta^{\mu}_{0}\delta^{0}_{\nu} - \rho_{\Lambda}\delta^{\mu}_{\nu}, \tag{4.14}$$

with $\rho_\Lambda = \Lambda/(8\pi G)$. We do so because the LTB metric can also be used to study the effect of inhomogeneities in an FRW universe [190].

The energy-momentum conservation equations $D_\mu T^{\mu}{}_{\nu} = 0$ are trivial for $\nu = (r, \theta, \phi)$. For $\nu = t$ it reads

$$\dot{\rho}_{\rm M} + (H_L + 2H_T)\,\rho_{\rm M} = 0 \tag{4.15}$$

Field equations

All the components of the curvature tensors of the LTB metric can be found in section C.2. Applying Einstein's equations ($G_{\mu\nu} = 8\pi G\, g_{\mu\sigma}T^{\sigma}{}_{\nu}$) to the metric and $T^{\mu}{}_{\nu}$ above we obtain two independent equations:

$$\frac{\left[A(\dot{A}^2 + k)\right]'}{A^2 A'} = 8\pi\, G\,(\rho_{\rm M} + \rho_\Lambda), \tag{4.16}$$

$$\frac{(A\dot{A}^2)^{\cdot}}{\dot{A}} + k = 8\pi\, G\,\rho_\Lambda\, A^2. \tag{4.17}$$

Multiplying (4.17) by $\dot{A}$ and integrating with respect to t we obtain a first integral

$$H_T^2 = \frac{8\pi\, G}{3}\,\rho_\Lambda - \frac{k}{A^2} + \frac{F(r)}{A^3}, \tag{4.18}$$

where $F(r)$ is an r-dependent function whose physical interpretation will be dealt with shortly. Substituting $\dot{A}^2 + k = \frac{8\pi G}{3}\rho_\Lambda A^2 + \frac{F}{A}$ into (4.16) we obtain

$$8\pi\, G\,\rho_{\rm M}(r,t) = \frac{F'(r)}{A^2(r,t)A'(r,t)}. \tag{4.19}$$

Evaluating this at $t = t_0$ and integrating over r

$$F(r) = 8\pi\, G \int_0^r \rho_{M,0}(r')\, r'^2\, dr' \equiv 2\, G\, M(< r), \tag{4.20}$$

thus $F(r)$ can be interpreted as being proportional to the amount of matter enclosed in a sphere of radius r.

As always, energy-momentum conservation (4.15) can be derived due to the Bianchi identities. In our case, we only have to differentiate (4.19) with respect to t. Summarizing, the equations governing the evolution of the metric and the density field are

$$H_T^2 = \frac{8\pi\, G}{3}\rho_\Lambda - \frac{k}{A^2} + \frac{F(r)}{A^3}, \quad 8\pi\, G\, \rho_{\rm M}(t,r) = \frac{F'(r)}{A^2(t,r)A'(t,r)} \tag{4.21}$$

The FRW analogy. Some physical interpretations

Let us define the functions $\Omega_{\rm M}(r)$ and $\Omega_\Lambda(r)$ as

$$F(r) \equiv H_0^2(r)\, \Omega_{\rm M}(r)\, A_0^3(r), \quad \Omega_\Lambda(r) \equiv \frac{8\pi\, G\, \rho_\Lambda}{3\, H_0^2(r)} \tag{4.22}$$

where $H_0 = H_T(t = t_0, r)$. Then we see that the evolution equation above can be written as

$$H_T^2(t,r) = H_0^2(r)\left[\Omega_{\rm M}(r)\left(\frac{A_0(r)}{A(t,r)}\right)^3 + \Omega_\Lambda(r) + \Omega_k(r)\left(\frac{A_0(r)}{A(t,r)}\right)^2\right], \tag{4.23}$$

where

$$\Omega_k(r) = 1 - \Omega_{\rm M}(r) - \Omega_\Lambda(r) = -\frac{k(r)}{H_0^2(r)A_0^2(r)}. \tag{4.24}$$

This is in direct analogy with Friedmann's equations for the FRW model, only that, in this case, the cosmological parameters have a radial dependence. Also, combining (4.16) and (4.17) we can derive the LTB generalization of the second Friedmann equation (1.5):

$$\frac{2}{3}\frac{\ddot{A}}{A} + \frac{1}{3}\frac{\ddot{A}'}{A'} = -\frac{4\pi G}{3}\rho_{\rm M} \tag{4.25}$$

It is also useful to work out a more appealing formula for the effective parameter $\Omega_{\rm M}(r)$. Notice that the first equation in (4.22) can be rewritten as

$$\Omega_{\rm M}(r) = \frac{\langle \rho_{\rm M}(r,t_0)\rangle_r}{\rho_{\rm crit}(r,t_0)}, \quad \text{where} \quad \langle f(r)\rangle_r \equiv \frac{\int_0^r f(r')\, r'^2\, dr'}{\int_0^r r'^2\, dr'}, \quad \rho_{\rm crit} \equiv \frac{3H_0^2(r)}{8\pi\, G}. \tag{4.26}$$

That is, $\Omega_{\rm M}(r)$ is the ratio of the mean density within a sphere of radius r and the critical density at this radius. Hence all the dynamics at a fixed r are determined by the amount of matter enclosed within this radius.

Analytic solution for voids

eq. (4.23) can be readily integrated to give the comoving time

$$t_0 - t(r) = \frac{1}{H_0(r)}\int_{\frac{A(r,t)}{A_0(r)}}^1 \frac{dx}{x\,\sqrt{\Omega_{\rm M}(r)\, x^{-3} + \Omega_k(r)\, x^{-2} + \Omega_\Lambda(r)}}. \tag{4.27}$$

Taking $A(r,t) \to 0$ one obtains the time to the Big-Bang.

In the most interesting case we study a DE-free ($\Lambda = 0$) void ($\Omega_{\rm M}(r) < 1$), and this last integral can be calculated explicitly [64]:

$$H_0(r)t_{BB}(r) = \frac{1}{\Omega_k(r)} - \frac{\Omega_{\rm M}(r)}{(\Omega_k(r))^{3/2}} \sinh^{-1}\sqrt{\frac{\Omega_k(r)}{\Omega_{\rm M}(r)}}. \tag{4.28}$$

Notice that since our model is not homogeneous nothing prevents the Big-Bang from occurring at different epochs in different locations. However, since we want to describe a spherical void that grows with time from a small perturbation at an early epoch (hence at this early epoch the Universe should be regarded as being homogeneous), and we want to prevent an unobserved isocurvature mode, we will be interested mainly in solutions with a homogeneous Big-Bang. Thus, for a void with a density profile given by $\Omega_{\rm M}(r)$, eq. (4.28) determines the present effective Hubble parameter profile $H_0(r)$:

$$H_0(r) = t_{BB}^{-1}\left[\frac{1}{\Omega_k(r)} - \frac{\Omega_{\rm M}(r)}{(\Omega_k(r))^{3/2}} \sinh^{-1}\sqrt{\frac{\Omega_k(r)}{\Omega_{\rm M}(r)}}\right], \tag{4.29}$$

where t_{BB} is the age of the Universe.

The $r-$dependent Friedmann equation (4.23) for a void ($\Omega_\Lambda = 0$, $\Omega_{\rm M} < 1$) can be solved analytically [64], and we can give an expression for $A(t,r)$ in terms of an additional parameter η,

$$A(t,r) = \frac{\Omega_{\rm M}(r)}{2[1-\Omega(r)]}\left[\cosh(\eta) - 1\right] A_0(r), \tag{4.30}$$

$$t = \frac{1}{H_0(r)} \frac{\Omega_{\rm M}(r)}{2[1-\Omega(r)]^{3/2}}\left[\sinh(\eta) - \eta\right]. \tag{4.31}$$

Given the functions $H_0(r)$ and $\Omega_{\rm M}(r)$ describing our void, for any (t,r), the corresponding value of η can be found by solving the second (transcendent) equation. Substituting this value in the first equation gives $A(t,r)$, which determines the metric.

4.1.5 Void models

In the previous sections we have shown that the dynamics of an inhomogeneous, spherically-symmetric universe, at a fixed radius r are, loosely speaking, those of an FRW universe with some $r-$dependent effective cosmological parameters $\Omega_{\rm M}(r)$, $\Omega_\Lambda(r)$ and $H_0(r)$. We have also shown that these effective parameters are determined by two free functions of r. We have found it convenient to take these two functions to be $H_0(r)$ and $\Omega_{\rm M}(r)$, since they have clear physical interpretations. Furthermore, if we are interested in models with a homogeneous Big-Bang, we have shown that this restriction completely determines the form of $H_0(r)$ in terms of $\Omega_{\rm M}(r)$, so we are left with just one free function.

Different groups have used different types of density profiles for their void models [191, 192, 193, 65]. Some use generic profiles showing desired features (such as compensating bumps at the void-background transition or a second central underdensity) and others simplistic top-hat density profiles. For our work we used the so-called GBH (García-Bellido-Haugbølle [64]) model, described below, in which the all the information about the density profile is encapsulated in a discrete set of parameters. This is very useful when performing a likelihood analysis against observational data.

The GBH model

In order to describe a void we need to find an expression for the free function $\Omega_{\rm M}(r)$ that satisfies the following requirements:

- We will consider models without dark energy that are, furthermore, asymptotically flat (so as not to spoil the predictions from inflation). Therefore we must require $\Omega_{\rm M}(r \to \infty) \to 1$.
- The central region (within a given radius r_0) should be devoid of matter, therefore $\Omega_{\rm M}(r < r_0) < 1$.

There are many ways to satisfy both requisites, for the GBH model one uses the following expression:

$$\Omega_{\rm M}(r) = 1 + (\Omega_{\rm in} - 1)\left[\frac{1 - \tanh[(r - r_0)/2\Delta r]}{1 + \tanh[r_0/2\Delta r]}\right], \tag{4.32}$$

where $\Omega_{\rm in} = \Omega_{\rm M}(0)$, r_0 is a measure of the radius of the void and Δr characterizes the width of the transition between void and background. This, together with the expression (4.29) for $H_0(r)$, completely determines the model. Thus we have reduced the degrees of freedom from 2 free functions of r to a set of 4 parameters: H_0, $\Omega_{\rm in}$, r_0, and Δr.

4.1.6 Perturbation theory in LTB

In homogeneous cosmological models one can find independent equations for scalar, vector and tensor perturbations, which simplifies the analysis, since one can focus on a given class of perturbations without having to solve for the other ones. This property is due to the fact that the background model is rotationally invariant at every point in space, and thus the irreducible representations of $SO(3)$ are uncoupled. This is no longer true in LTB, since the model is only $SO(3)$-invariant at $r = 0$. The only way to obtain decoupled perturbation equations in LTB is to find the irreducible representations of the remaining symmetry group, rotations that use the line of sight as an axis. This has been done in [194] and, even though the resulting set of perturbation equations turns out to be remarkably complicated, certain solutions can be found and predictions have been given for the shape of the radial and angular two-point correlation functions in real space [195]. This is an important step forward in order to address problems like the study of the BAO in LTB models, however more work is needed, like introducing redshift-space distortions and studying higher-order corrections.

If, notwithstanding what has just been said, we try to find the perturbation equations for scalar, vectors and tensors, we find that, in particular, the scalar modes couple to the vector and tensor modes through the background shear tensor $\Sigma_{\mu\nu} \propto H_T - H_L$, which is non-zero in LTB models, and that in the limit $\Sigma_{\mu\nu} \to 0$, the standard FRW perturbation equations are recovered. The amount of shear associated to a given void model can be quantified in terms of the shear-to-expansion ratio

$$\epsilon = \frac{H_T - H_L}{2\,H_T + H_L}. \tag{4.33}$$

See section 4.2.2 and eq. (4.43) for a more explanatory derivation of this quantity.

In viable LTB models, the background shear is relatively small ($\epsilon \lesssim 0.1$) and thus we can assume that all shear-dependent terms in the perturbation equations are subdominant. In this approximation the equation for the scalar Newtonian potential becomes

$$\ddot{\Phi}(r,t) + 4\,H_T(r,t)\,\dot{\Phi}(r,t) - \frac{2k(r)}{A^2(r,t)}\,\Phi(r,t) = 0\,, \tag{4.34}$$

which, as we said, is the FRW equation with $H(t) \to H_T(r,t)$. The exact solution (growing mode) [72] can be written in terms of the density contrast as

$$\delta_0(t,r) = B\, D(\Omega_{\rm M}(r), A(t,r)/A_0(r)), \tag{4.35}$$

where $D(\Omega_{\rm M}, a)$ is the growth factor in an FRW model with matter parameter $\Omega_{\rm M}$, given by

$$\begin{aligned} D(\Omega_{\rm M}, a) &= \frac{5}{2}\,\Omega_{\rm M}\, h^2\, H(\Omega_{\rm M}, a) \int_0^a \frac{da'}{[a'\, H(\Omega_{\rm M}, a')]^3} \\ &= a \cdot {}_2F_1\left[1, 2; \frac{7}{2}; \frac{\Omega_{\rm M}-1}{\Omega_{\rm M}}\, a\right]. \end{aligned} \tag{4.36}$$

(Here ${}_2F_1[a,b;c;z]$ is the Gauss hypergeometric function).

We reiterate that, strictly speaking, this solution is only exact when ignoring the tensor coupling, and considering angular transverse modes, but, as we will see in section 4.3.4 turns out to be a very good approximation. The proportionality factor B in (4.35) depends on r, but will be irrelevant for our purposes, since we will normalize the growth factor to $\delta(t_0, r) \equiv 1$. Nonetheless, for completeness, it is possible to find $B(r)$ up to a constant factor under the assumption that the small-scale matter perturbations in the early Universe decouple from the void, giving $B(r) \propto r/A(r, t_{\rm early})$.

Since the background shear is small in the models we have explored, we can parametrize its effect on the growth of perturbations by a series expansion in ϵ, and to linear order we can write

$$\delta_\alpha(t,r) = \delta_0(t,r)\,(1 + \alpha\,\epsilon(t,r))\,, \tag{4.37}$$

where the parameter α might in principle depend on (t,r) and also on the cosmological model.

4.2 Observational tests of LTB models

LTB models have been tested against many cosmological observations. We will describe them and their associated constraints here together with those that future experiments could impose with newly proposed observables.

4.2.1 Contrast with present observations

- **Type Ia supernovae.** Observations of SNeIa gave the first hint of an apparent cosmic acceleration and, as alternative models to dark energy, void models are always subjected to this test first. The distance-redshift relation is probed through the distance modulus $\mu \equiv m - M$, where m and M are the object's relative and absolute magnitudes. μ is related to the luminosity distance through

$$\mu(z) = 5\,\log_{10}\left(\frac{d_L(z)}{1\,{\rm Mpc}}\right) + 25. \tag{4.38}$$

  Several groups have contrasted the predictions of LTB void models with SNeIa observations [198, 64, 196, 65, 199, 200, 191, 192, 201], see fig. 4.1). In general the conclusion is that one can always find a void profile (even a generic parametrized one like the GBH model) that yields a distance-redshift relation in accordance with observations and with the concordance ΛCDM model. Therefore other observables are necessary to distinguish them.

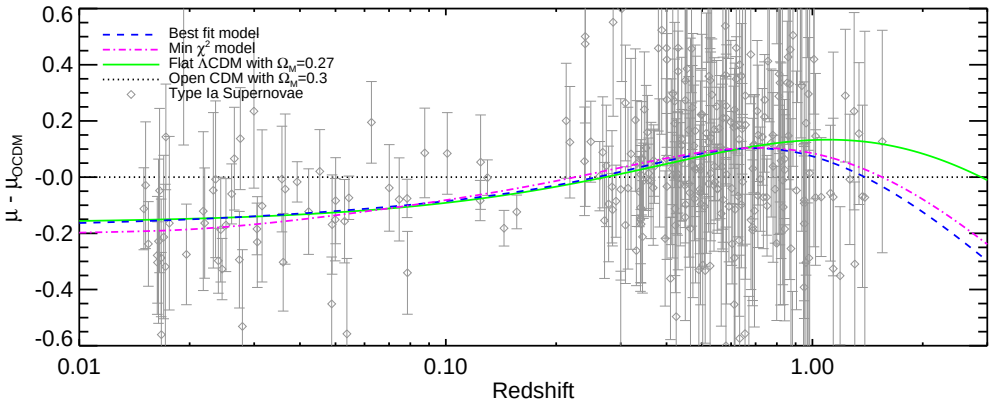

Figure 4.1: Distance modulus from 307 SNeIa from the UNION catalogue together with the predictions for a best-fit LTB model and the concordance ΛCDM model. Figure taken from [196].

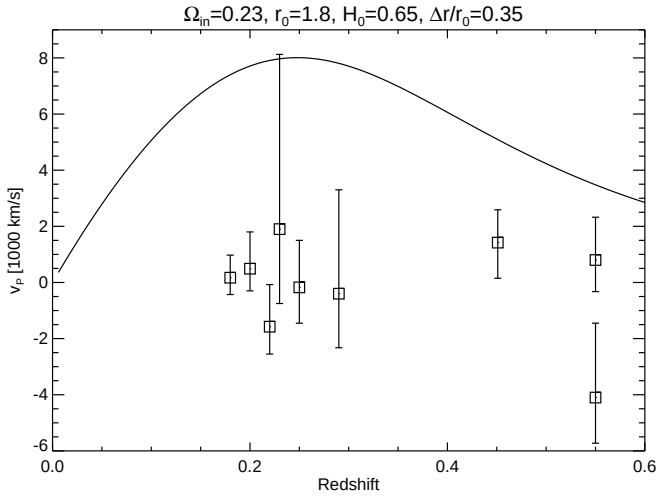

Figure 4.2: Peculiar velocities for 9 galaxy clusters obtained from measurements of the kSZ effect together with the LTB velocity profile for a best-fit void. Although the data are few they already present a challenge to void models. Figure taken from [197].

- **BAO.** As we have explained, the position of the baryon acoustic peak in the two-point correlation function of matter perturbations can be used as a standard ruler to study the distance-redshift relationship. BAO constraints coming from the study of the monopole of the 2PCF are usually given in terms of the ratio

$$\frac{r_s}{D_V(z)}, \tag{4.39}$$

where r_s is the comoving sound horizon at recombination and D_V is a geometric mean of comoving radial and transverse distances [202]:

$$D_V(z) \equiv \left(d_A^2(z)\,(1+z)^2 \frac{c\,z}{H(z)} \right)^{1/3}. \tag{4.40}$$

BAO measurements have been compared with LTB predictions [64, 191, 192], but only lately [201] it has been shown that the combination SNe and BAO data impose very severe constraints on LTB models with a homogeneous Big-Bang.

BAO measurements can also be performed in the angular and radial directions separately. The advantages and drawbacks of this approach are described in chapter 3. The main gain in our case is the fact that these two measurements allow us to disentangle the radial and transverse evolution, which are intrinsically different in LTB cosmologies. In the case of radial BAO measurements, statistical uncertainties are extremely large and so far only mild constraints can be obtained from them. The comparison of the few independent measurements that exist with LTB predictions [196, 199] shows that they favour larger voids ($r_0 \sim 2.7$ Gpc) than previous observables. Angular BAO measurements [29] have also been included in the joint constraints on LTB models in [201].

- **CMB spectrum.** As we have seen, Perturbation theory in LTB backgrounds is very complex and no exact predictions for the matter power spectrum in LTB models exist. This is discouraging and could in principle prevent any comparison with CMB data. However the following approximation can be used: the CMB spectrum depends mainly on the primordial spectrum of fluctuations, the epoch of matter-radiation equality and the dynamics of the baryon-photon fluid at recombination. Since the last scattering surface is beyond the void walls, in the FRW background and at early times the void is just a perturbation of small amplitude in this background, we only need to know the present energy densities of the various components (baryons, CDM, radiation, etc.) in the FRW background (this is subtle, since the density of radiation is computed from the CMB temperature, which is not the same in the interior of the void). Once this is done one can use a FRW Boltzmann code (e.g. `CAMB` [88]) to recreate the epochs of equality and recombination. Finally all scales at the CMB epoch must be evolved to the angular scales observed today through the distance-redshift relation. LTB predictions in this approximation have been compared with the CMB spectrum data by different groups [65, 199, 64, 196, 192, 203, 204, 201], with similar conclusions: LTB models which are asymptotically flat can fit simultaneously SNe and CMB data, however the inclusion of the latter favours models with a very low local Hubble parameter ($H_0/100 \simeq 0.5\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$). However, this problem can be alleviated by allowing the background cosmology to have a non-zero curvature, or by using a (physically motivated) non-standard primordial power spectrum [204]. It has also been noted [192] that if the void only reaches the background cosmology asymptotically the full treatment of radiation in an LTB metric must be performed in order to obtain reliable predictions.

- **Local expansion rate H_0.** Observationally the local Hubble parameter H_0 is defined as

$$H_0^{-1} \equiv \lim_{z\to 0} \frac{d_A(z)}{z} = H_T(r,t)|_{z=0}\,. \tag{4.41}$$

  It has been measured by the HST (Hubble Space Telescope) from low redshift supernovae to be $H_0 = 62.3 \pm 6.3\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ [205]. More recent and accurate measurement using the Cepheid distance scale push this value even higher: $H_0 \simeq 73 - 74\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ [206, 207]. As we mentioned, CMB data favours asymptotically flat models with very low H_0, which are in conflict with this value. Note that [192] and [199] include these measurements in the Monte-Carlo analysis differently, which is apparently the root of their different conclusions.

- **Kinetic Sunyaev-Zel'dovich effect (kSZ).** When relative motions between free electrons and CMB photons exist, the inverse Compton scattering induces a shift of the brightness temperature of the CMB, via the kinetic Sunyaev-Zel'dovich effect [208]. This effect has been used mainly to determine the peculiar velocities of galaxy clusters, through the kSZ effect induced by their hot gas, however it can be applied to all free electrons, which exist abundantly up to $z \sim 6$ (unlike clusters, which are rare above $z \sim 1$). The temperature anisotropy in the direction $\mathbf{n}$ due to this effect is

$$\frac{\Delta T(\mathbf{n})}{T} = \int \mathbf{v}(\mathbf{n},z)\cdot\mathbf{n}\,[1+\delta_e(\mathbf{n},z)]\,d\tau, \tag{4.42}$$

  where $\mathbf{v}$ is the velocity of free electrons relative to the CMB rest frame, $\delta_e = \delta\rho_e/\rho_e$ is the density contrast of free electrons and τ is the optical depth along the line of sight. It is crucial to note here that, unlike in the FRW case, in LTB models there exists a systematic "outflow" velocity with respect to the CMB frame, since comoving matter elements within the void move with respect to the background FRW. This is not a peculiar velocity, but a background effect, which makes the kSZ effect a potentially constraining tool. In other words: in FRW the dominant term in (4.42) is quadratic in the perturbations, since the velocities are peculiar (the term $\propto \int \mathbf{v}\cdot\mathbf{n}\,d\tau$ is negligible due to geometrical cancellation or contributes to an unobservable monopole and only the term $\propto \int \mathbf{v}\cdot\mathbf{n}\,\delta_e\,d\tau$ remains), however it is linear in LTB, since the average matter frame moves with respect to the CMB inside the void.

  LTB models have been tested against observations of the kSZ effect induced by clusters [197, 209] and by all free electrons [210, 211] with qualitatively equivalent results: in general LTB void models predict too large a kSZ contribution and these works find it quite difficult to accommodate the current data (see fig. 4.2). However there still exist theoretical ambiguities and observational shortcomings. In the case of clusters the existing data are still very few.

- **Other observables.** LTB predictions have been compared with other observables: Big-Bang Nucleosynthesis [199, 191], Compton y-distortion, age of the Universe [199] and LSS-related observables, such as σ_8 [199] or the matter power-spectrum [192]. These last two have very little constraining power at this stage, since LTB perturbation theory is still not fully understood.

4.2.2 Proposed observables and future experiments

As we said, observations in the near future will be able to probe the validity of some of the assumptions behind the standard cosmological model, namely the homogeneity and isotropy on large scales. To this end several observables have been proposed:

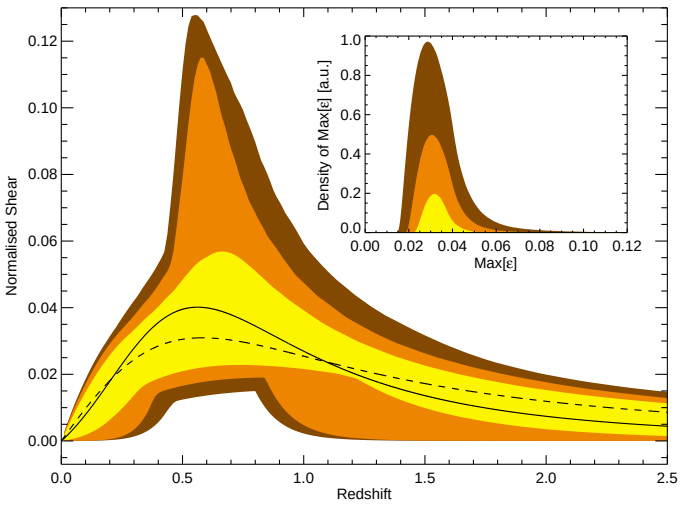

Figure 4.3: Normalized shear $\epsilon(z)$ for best-fit LTB models. Figure taken from [197]

- **Cosmic shear.** Raychaudhuri's theory [212] studies the way in which a congruence of freely falling test particles evolves in a given gravitational background. This evolution is described in terms of three properties of the congruence, defined from the tensor $B_{\mu\nu} \equiv D_\nu u_\mu$: the expansion scalar $\theta \equiv B^\mu_\mu$, describing the growth of the overall congruence volume, the shear tensor $\sigma_{\mu\nu} \equiv \frac{1}{2}P^\alpha_\mu P^\beta_\nu (B_{\alpha\beta} + B_{\beta\alpha}) - \frac{1}{3}\theta\, P_{\mu\nu}$, describing the distortion of the congruence shape without changing the volume, and the vorticity tensor $\omega_{\mu\nu} \equiv \frac{1}{2}P^\alpha_\mu P^\beta_\nu (B_{\beta\alpha} - B_{\alpha\beta})$, describing the rotation of the congruence. Here $P^\mu_\mu \equiv g^\mu_\nu + u^\mu u_\nu$ is a projector onto the plane orthogonal to the congruence (see [213] for more details on this formalism). Using this approach to study congruences of comoving observers in FRW and LTB cosmologies one can realise that they only vary in volume in FRW ($\sigma_{\mu\nu} = 0, \omega_{\mu\nu} = 0$), while they also vary in shape for a general LTB metric ($\sigma_{\mu\nu} \neq 0, \omega_{\mu\nu} = 0$), which is mainly due to the fact that the expansion rates in the radial and transverse directions are different, and therefore tidal forces tend to distort the shape of extended structures. This motivates the definition of the normalized shear-to-expansion ratio [196]

$$\epsilon \equiv \sqrt{\frac{3}{2}}\frac{\sqrt{\sigma_{\mu\nu}\,\sigma^{\mu\nu}}}{\theta} = \frac{H_T - H_L}{H_L + 2H_T} \tag{4.43}$$

This cosmic shear can be expressed in terms of observable quantities

$$\epsilon(z) \simeq \frac{1 - H_L(z)\,\partial_z[(1+z)\,d_A(z)]}{3\,H_L(z)\,d_A(z) + 2\,(1 - H_L(z)\,\partial_z\,[(1+z)\,d_A(z)]).} \tag{4.44}$$

A detection of a non-zero ϵ would be a clear signature of an inhomogeneous expansion ($\epsilon(z)$ can be of the order of ~ 0.1 for viable LTB void models, see fig. 4.3), but, as is evident from eq. (4.44) and as we had already mentioned, independent measurements of radial and angular distance-redshift relations are needed.

- **Cosmic parallax.** In a homogeneous universe, or in a universe with spherical symmetry around the observer, and assuming that peculiar velocities are 0, the angle subtended by two distant objects would always be constant. However, if the expansion is not perfectly isotropic (as would be the case if the observer were slightly off-centred in an

LTB void), this angle will vary. Measuring the time variation of this angle has been proposed [214, 215, 216] as a general test of the isotropy of the Universe. Consider two observers: C, located at the centre of spherical symmetry, and O, at a distance d_O from C. Angles measured by both with respect to the C-O axis for an object at a distance d from C are related by

$$\cos\theta_O = \frac{d\,\cos\theta_C - d_O}{\sqrt{d^2 + d_O^2 - 2\,d_O\,d\,\cos\theta_C}}. \tag{4.45}$$

Now consider two sources (in the same plane as C and O and at the same distance d from O for simplicity), which have an angular separation γ observed by O. If the expansion is homogeneous d and d_O will vary after a time Δt: $\Delta d = d\,H(t)\,\Delta t$, $\Delta d_O = d_0\,H(t)\,\Delta t$. Substituting in (4.45) we see that all the $1 + H(t)\,\Delta t$ factors cancel out and γ does not vary, however in LTB $\Delta d = d\,H_L(r,t)\,\Delta t$, $\Delta d_O = d_O\,H_L(r_O,t)\,\Delta t$ and no cancellation occurs. It can be proven [214] that for two sources separated by a distance Δs and subtending an angle $\Delta\theta_C$ observed by C, the variation of the parallax observed by O is

$$\begin{aligned}\Delta\gamma =& \frac{d_O}{d}\,\Delta t\,(H_L(t_0,r_O) - H_L(t_0,r))\left(\cos\theta\,\Delta\theta_C + \sin\theta\,\frac{\Delta s}{d}\right)\\ &+ H_L(t_0,r)\frac{dH_L(t_0,r)}{dr}\sin\theta\,\Delta s + \mathcal{O}\left(\frac{d_O}{d}\right)^2.\end{aligned} \tag{4.46}$$

This variation is non-zero in any cosmology with a non-homogeneous expansion and its detection would allow us to probe the validity of the CP as well as to constrain our position with respect to the centre of the void in an LTB model. Of course, peculiar velocities will also produce a variation in the parallax, and this should be the main source of uncertainty, however this effect should average out when measured for a sufficiently large set of sources. In [214] a measurement of $\Delta\gamma = 10\frac{d_0}{d}\,\mu$arc.sec is predicted over a period of 10 years of observations for some future space missions such as Gaia [217].

The measurement of the cosmic parallax, together with the redshift drift, which we describe below, belong to a new field of research called *real-time cosmology* [218], based on consecutive astronomical observations of the same objects separated by a few years.

- **Redshift drift.** For any expanding cosmology the redshift of a comoving object is not constant in time, and its variation could be an interesting observable to constrain dark energy models. For decelerating expansion the redshift drift $\dot{z}$ is always negative, however in models with late-time acceleration, like ΛCDM or quintessence models, sources at $z \lesssim 2$ have growing redshifts. Thus, alternative models with only virtual acceleration, such as LTB models, have $\dot{z} < 0$. This effect on void models has been studied by several groups [219, 220, 216] and could in principle be measured by future ELTs (Extremely Large Telescopes).

- **Gravitational lensing.** The bending of the light from a distant source due to the distribution of matter along its trajectory is known as gravitational lensing (GL). The mathematical formalism to study gravitational lensing in general spherically symmetric spaces has been developed [221] and predictions about its effects on the spectra of the E and B modes of the CMB for LTB cosmologies have been extracted [222].

4.2.3 Fine-tuning in void models

The observations described in the previous sections constrain the properties of viable void models. Considering only asymptotically flat LTB models, the preferred voids have a radius

of $r_0 \sim 2-3$ Gpc, a central underdensity of $\Omega_{\rm in} \sim 0.2$ and a central expansion rate of $H(r=0) \sim 0.6$ km s^{-1} Mpc^{-1} [196]. Let us reflect on the naturality of this result.

First there is the issue of the probability that such a huge perturbation exists within our observable universe. This issue has been addressed within the theory of eternal inflation [223], in which rare fluctuations at the Plank boundary may be responsible for the non-perturbative amplification of local inhomogeneities in the metric, which would look like local voids in the matter distribution. In fact such a void might have already been observed as the CMB cold spot [224], which could have properties compatible with the ones above.

Then there is the problem of the spatial fine tuning: observations suggest that if indeed we reside in a Gpc-sized void, it must be spherically symmetric and we must live close to its centre, otherwise we would observe a large dipole component in the CMB anisotropies. This dipole is indeed observed but it is assumed to be caused by the presence of the Virgo cluster and the Shapley super cluster. Taking all this into account and neglecting the possibility that our peculiar velocity could point towards the void centre (which would partially cancel this dipole) one can set bounds on our possible offset [193, 225], which cannot be larger than about ~ 100 Mpc. This, compared with the much larger size of the void could present a probabilistic challenge to LTB models. However, again in the eternal inflation approach, the advocated rare fluctuations should be highly spherically symmetric and typical observers should live close to their centre [226].

Finally, it has been noted that LTB models only behave like a virtual ΛCDM universe for some period of time, and therefore also suffer from a coincidence problem. A thorough treatment of fine-tuning issues in LTB models can be found in [225].

4.3 Large-scale structure simulations of LTB models

Even though their motivation, a reinterpretation of the distance-redshift relation measurements without any need for DE, is very attractive, LTB models suffer from many flaws. The most evident one is the violation of one of the most fundamental philosophical principles: the Copernican Principle. The most important one, however, is their failure to match several cosmological observations, e.g.: the combination of CMB and H_0 data, the tension between BAO and SNeIa and the amplitude of the kSZ effect. The contrast of LTB models with observational data has put very tight constraints on LTB models, ruling out the most natural ones. However not all of the constraining power of the current datasets has been exploited. This is due to the theoretical uncertainties of LTB models: our understanding of perturbation theory and structure formation is very poor and far behind the state of the art in homogeneous models. A more thorough understanding of these problems would allow us to constrain LTB models further by comparing their predictions with LSS data from galaxy redshift surveys or gravitational lensing experiments.

As we know, structure formation occurs via gravitational collapse, and is a non-linear process. Even in the framework of the standard cosmological model, where perturbation theory is well understood, the full non-linear dynamics can only be reliably studied in a numerical way through N-body simulations. Hence, even if we had a complete understanding of PT in LTB models, structure formation must ultimately be studied through these simulations.

In this section we will describe the study that motivated the publication of the paper [7]. The idea was to use an N-body simulation to follow the evolution of a Gpc-sized void. This is interesting for different reasons. First of all, this is the first time that a cosmological N-body code has been used to simulate an inhomogeneous cosmological model. Devising a technique to do so and verifying that the theoretical model (at least at the background level) is well

reproduced is interesting in itself. Furthermore, when refined, this technique would open the possibility of studying LSS on an LTB background at the non-linear level.

In section 4.3.1 we describe the technique used to simulate the Gpc voids. In 4.3.2 we provide an overview of the different simulations performed. Then in 4.3.3 we present the results obtained for the background evolution. We have also attempted a first study of structure formation in LTB models using N-body simulations, which is described in section 4.3.4. Finally we present our conclusions in section 4.3.5.

4.3.1 Simulating voids

Our approach to simulate LTB models is to understand the void as a perturbation within a background FRW universe. This interpretation is indeed perfectly accurate, since at early times the void is a perturbation of very small amplitude ($\sim 10^{-3}$ at $z \sim 1000$), albeit large in scale, that follows a non-linear evolution via gravitational collapse. Perturbations in N-body simulations are generated by displacing the dark matter particles from their comoving positions in the initial conditions, therefore we only need to determine the additional displacement due to the presence of the void. To do this we modified the `2LPT` [99] initial condition generator, which uses second-order LPT, in the following way:

1. We calculate a standard transfer function for the perturbations in the background FRW model and `2LPT` generates a realization of the initial gravitational potential in Fourier space $\Phi^p_{\mathbf{k}}$.

2. The gravitational potential of a void $\Phi^v_{\mathbf{k}}$ is found using the analytical solution at the starting redshift, by interpolating the density out on the particle grid and then Fourier-transforming it.

3. The particle displacements are calculated from the total potential $\Phi_{\mathbf{k}} = \Phi^p_{\mathbf{k}} + \Phi^v_{\mathbf{k}}$ by `2LPT` using Lagrangian perturbation theory.

Once the initial conditions are set we feed them to the N-body code `GADGET2` [126], which will evolve the system through its full Newtonian dynamics in an expanding background given by the background FRW model. For our simulations we used an Einstein-de Sitter model with parameters $(\Omega_{\mathrm{M}}, f_{\mathrm{b}}, h, \sigma_8, n_s) = (1, 0.14, 0.43, 0.9, 1)$.

Initially a few simulations were performed using the faster TreePM version of `GADGET2`, however we realized that the mesh seemed to introduce an orthogonal pattern on top of the spherical symmetry of the model (see fig. 4.4). For safety and to avoid additional complications, all subsequent simulations, for which we present results, were run in pure Tree mode.

4.3.2 The simulations

It is not evident that N-body simulations can be used to describe large-scale LTB models, and therefore a significant effort has gone into validating that indeed we reproduce the expected theoretical behaviour. Several simulations have been run (see table 4.1 for an overview of all of them) to test the effects of varying different simulation parameters.

- We have used different starting redshifts ($z_{\mathrm{start}} = 24$, 49, 99 and 199) to check that the code is started at high enough redshift, such that the displacements of the particles are much smaller than the inter-particle distance, and that the void can be treated as a linear perturbation, which at first order does not interact with the small scale fluctuations from the power spectrum.

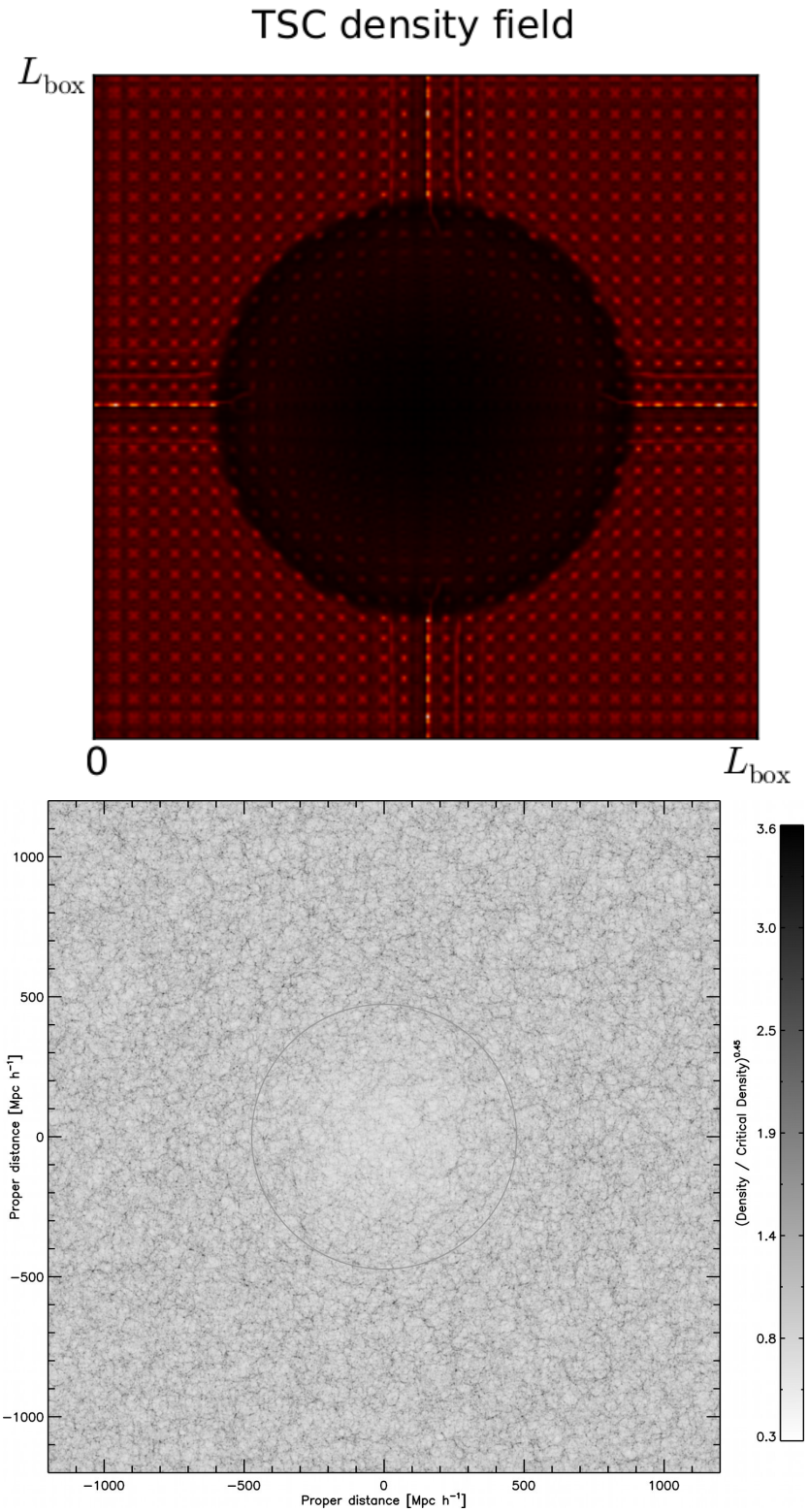

Figure 4.4: Top panel: 10 Mpc thick slice centred at the void position for a simulation run with the TreePM algorithm. Notice the orthogonal periodic features artificially created by the PM part and the effect of the box walls on the spherical void. Bottom panel: a similar slice for the high resolution simulation $\mathcal{H}$ can be seen. All the artificial features have been taken care of by running the code in pure Tree mode and enlarging the box.

Name	$z_{\rm start}$	$\Omega_{\rm in}$	$\Delta r/r_0$	#particles	Comments
$\mathcal{H}$	24	0.25	0.3	960^3	High res sim
$\mathcal{V}$	49	0.25	0.3	512^3	Void alone
$\mathcal{S}24$	24	0.25	0.3	512^3	Void + matter
$\mathcal{S}49$	49	0.25	0.3	512^3	Void + matter
$\mathcal{S}99$	99	0.25	0.3	512^3	Void + matter
$\mathcal{S}\Omega125$	49	0.125	0.3	512^3	Void + matter
$\mathcal{S}\Omega063$	49	0.0625	0.3	512^3	Void + matter
$\mathcal{S}\Omega021$	199	0.0208	0.3	512^3	Void + matter
$\mathcal{S}\Delta01$	49	0.125	0.1	512^3	Void + matter
$\mathcal{S}\Delta05$	49	0.125	0.5	512^3	Void + matter
$\mathcal{L}$	49	0.25	0.3	768^3	L=3600 Mpc h^{-1}

Table 4.1: Overview of the simulations. All have been performed with a void of radius $r_0 = 1100\,\mathrm{Mpc} = 473\,\mathrm{Mpc}\,h^{-1}$, and with an asymptotic Hubble parameter $h_\infty = 0.43$. The standard box size is $L = 2400\,\mathrm{Mpc}\,h^{-1}$, and the particle mass is $m_p = 2.8 \times 10^{13} M_\odot h^{-1}$ ($m_p = 4.3 \times 10^{12} M_\odot h^{-1}$ for $\mathcal{H}$). Everywhere we have used a smoothing length of 56 kpc h^{-1} (except for $\mathcal{H}$, where it has been appropriately rescaled).

- We have used different mass resolutions to test that the cosmological large-scale structure is adequately resolved. This was not our main concern, since our primary aim was to reproduce the background evolution, however it was interesting to see how far we could go with our computational resources. The best mass resolution we achieved, $m_p = 4.3 \times 10^{12}\, M_\odot h^{-1}$ is still a bit far from a desirable value $\mathcal{O}(10^{10-11} M_\odot h^{-1})$, but it is small enough to make a first attempt at studying LSS in LTB models (see section 4.3.4).

- We checked that to first order the small-scale fluctuations do not back-react significantly on the void by running analogous versions of the same void with and without matter perturbations.

- Since the simulation box has periodic boundaries, effectively we are not simulating one single void, but an infinite network of voids. We have tested that the void does not interact too much with mirror images of itself by changing the physical box size from $L = 2.4$ to $L = 3.6\,\mathrm{Gpc}\,h^{-1}$. We observed that several defects were induced for inadequate box sizes, including orthogonal patterns (see figure 4.4, top panel) and an unphysical cusp at about $r = r_0$.

Our most sophisticated simulation ($\mathcal{H}$) was performed for a fiducial GBH void model with $\Omega_{\rm in} = 0.25$, $H_0/100 = 0.64\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Delta r/r_0 = 0.3$. We varied these parameters in the other simulations in order to probe the limits of the validity of the N-body approach. In particular we also used $\Omega_{\rm in} = 0.125$, 0.625 and 0.0208, and $\Delta r/r_0 = 0.1$ and 0.5.

4.3.3 Background evolution

In order to verify the agreement between the simulations and the analytical solutions, we used the fiducial simulation $\mathcal{H}$ as our reference model and the other simulations to test the limits of this agreement. To do this we have studied the density and velocity profiles (the latter in terms of the expansion rates H_T and H_L).

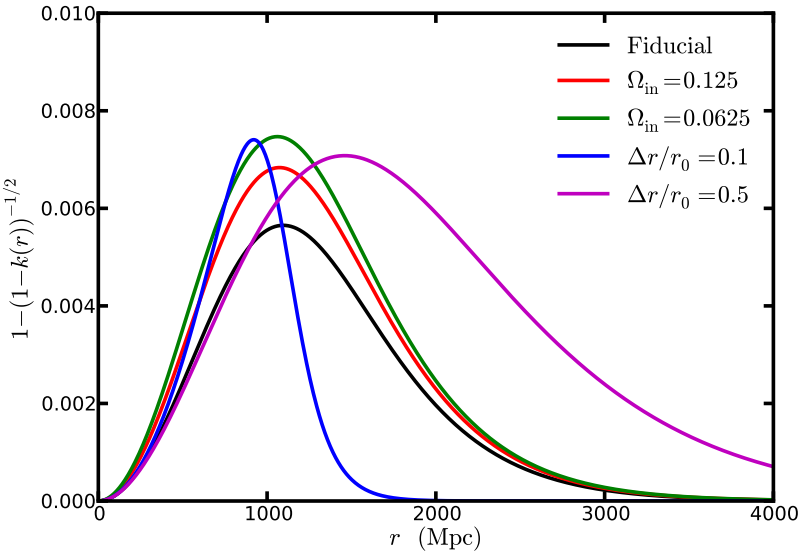

Figure 4.5: $(1-k(r))^{-1/2}$ for the models under consideration. In all cases they differ from 1 in less than 1%.

Distances and redshifts

GADGET2 is designed to perform simulations of FRW universes, and one needs to associate the comoving coordinates used by the code with the ones in the LTB model. Since we have analyzed the data from GADGET2 snapshots, that is, positions and velocities of particles are "measured" at constant cosmic time, and all our observables are quantities calculated in thin spherical shells, this identification must be done through the proper radial distance, calculated in both cases as

$$d_p(r,t) = a(t)\, r_{\rm FRW} = \int_0^{r_{\rm LTB}} \frac{A'(r,t)}{\sqrt{1-k(r)}}\, dr\,. \tag{4.47}$$

If the curvature factor $(1-k(r))^{-1/2}$ is roughly 1, which is the case in the models under study (see fig. 4.5), one can approximate

$$d_p(r,t) = a(t)\, r_{\rm FRW} \simeq A(r_{\rm LTB}, t) \tag{4.48}$$

for most redshifts. Similarly, when interpreting the results, it is important to remember that while the proper cosmological time in the two metrics can readily be identified, the redshifts at equal times are different, i.e. for $t_{\rm FRW} = t_{\rm LTB}$ the $z_{\rm FRW}$ and $z_{\rm LTB}$ are different. It is important to emphasize that since we are considering a constrained-GBH LTB model, the time to Big Bang is homogeneous and thus all times at each radial comoving distance are the same, so each particle in the simulation has a time given by the code: $t_{\rm FRW} = t_{\rm LTB}$.

Density profiles

In order to compare the density profiles in the simulations with their expected values, we must first estimate the density field from the particle positions, for which there exist different prescriptions. The idea is to interpolate the particle content of the box into a lattice using some algorithm. The naive way of doing it would be to just count the number of particles in each cell and divide by the cell volume (this is the nearest-grid point scheme or NGP).

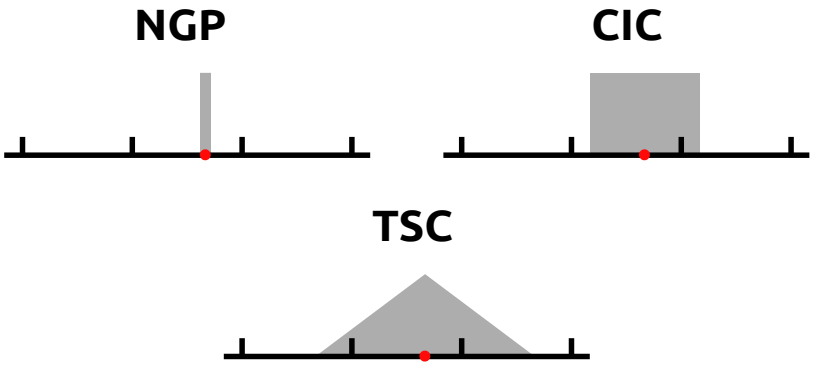

Figure 4.6: 1-dimensional illustration of the NGP, CIC and TSC algorithms to interpolate the particle content of a simulation into a density grid. The weight of the particle's contribution to each cell is proportional to the area of the part of the shaded region lying within the cell.

More sophisticated methods involve giving a weight to the cells close to each particle based on the distance from the particle to the cell. We have used a 2^{nd} order triangular-shaped-cloud (TSC) technique (see fig. 4.6). Once this is done, the simulation box is divided into different spherical shells, and we calculate the average density in each of them thus obtaining the density as a function of the proper distance d_p.

Due to the presence of non-linear inhomogeneities, the error in the determination of the density profile cannot be directly obtained as the r.m.s. in each bin, and the error bars displayed in the figures have been calculated as the r.m.s. in the analogous $\mathcal{V}$ simulation without CDM perturbations. The reference simulation $\mathcal{H}$ shows an excellent agreement between theory and simulation (see fig. 4.7), except near the centre of the void, where the particle distribution is undersampled and shot noise dominated.

In fig. 4.8 we show the density profile for an extended set of models. For most models the simulations are in excellent agreement with the theory, although for two extremal cases, namely the emptiest void $\mathcal{S}\Omega021$, and the void with the steepest transition $\mathcal{S}\Delta01$ we find significant deviations. For $\mathcal{S}\Delta01$ the discrepancy is not severe, and only present in the density profile. We speculate that this could be due to under-resolution of the transition length or possibly due to the small-scale perturbations interacting with the large-scale void, given that the transition length is only $\Delta r = 47.3\,\text{Mpc}\,h^{-1}$.

Expansion rate

The radial velocity profile can be used to compare against the theoretical predictions for H_T and H_L. The rate of change in the proper distance $\dot{d}_p/d_p$ computed in the rest-frame of the matter should match each other in the FRW and LTB metric, if the simulations are a valid description of the LTB model. In the LTB metric, unperturbed matter is at rest and keeps the same comoving coordinate, while in the FRW frame there are systematic radial motions, and we have that

$$\frac{d}{dt} d_p^{\text{FRW}} = \frac{d}{dt}\left[a\, r_{\text{matter}}\right] = d_p\left[\langle v_r\rangle / r + H_\infty\right] , \tag{4.49}$$

which can be directly compared to the theoretical LTB result, calculated by taking the derivative of the r.h.s. of eq. (4.47). $\langle v_r\rangle$ is calculated as the average radial velocity v_r of the particles sampled in spherical bins. In the upper panel of fig. 4.9 we see how the theoretical radial

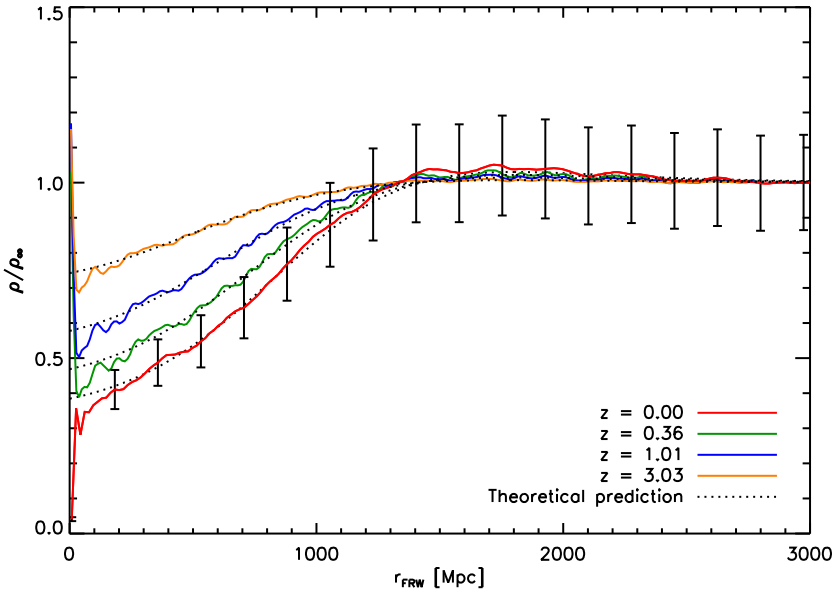

Figure 4.7: Comparison of the density profile of the $\mathcal{H}$ simulation at different redshifts with the theoretical curves, as a function of comoving distance $r_{\mathrm{FRW}} = (1+z)d_p$ in Mpc.

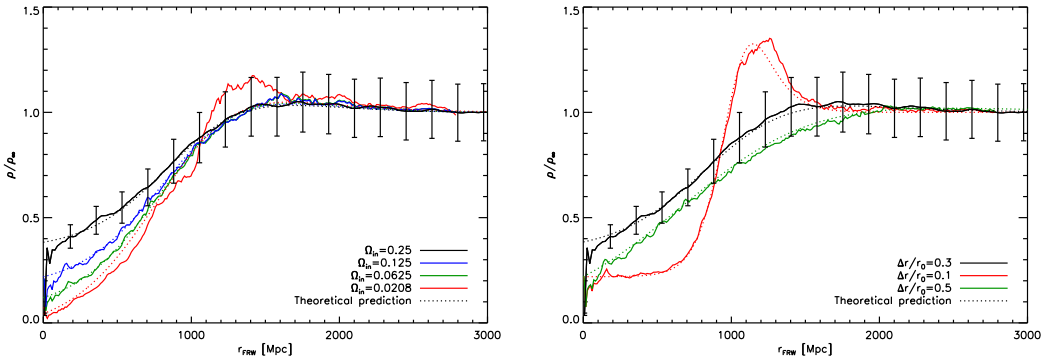

Figure 4.8: Density profiles for different values of Ω_{in} (left panel) and $\Delta r/r_0$ (right panel) in comparison with the corresponding theoretical profiles. All curves are plotted at redshift $z = 0$.

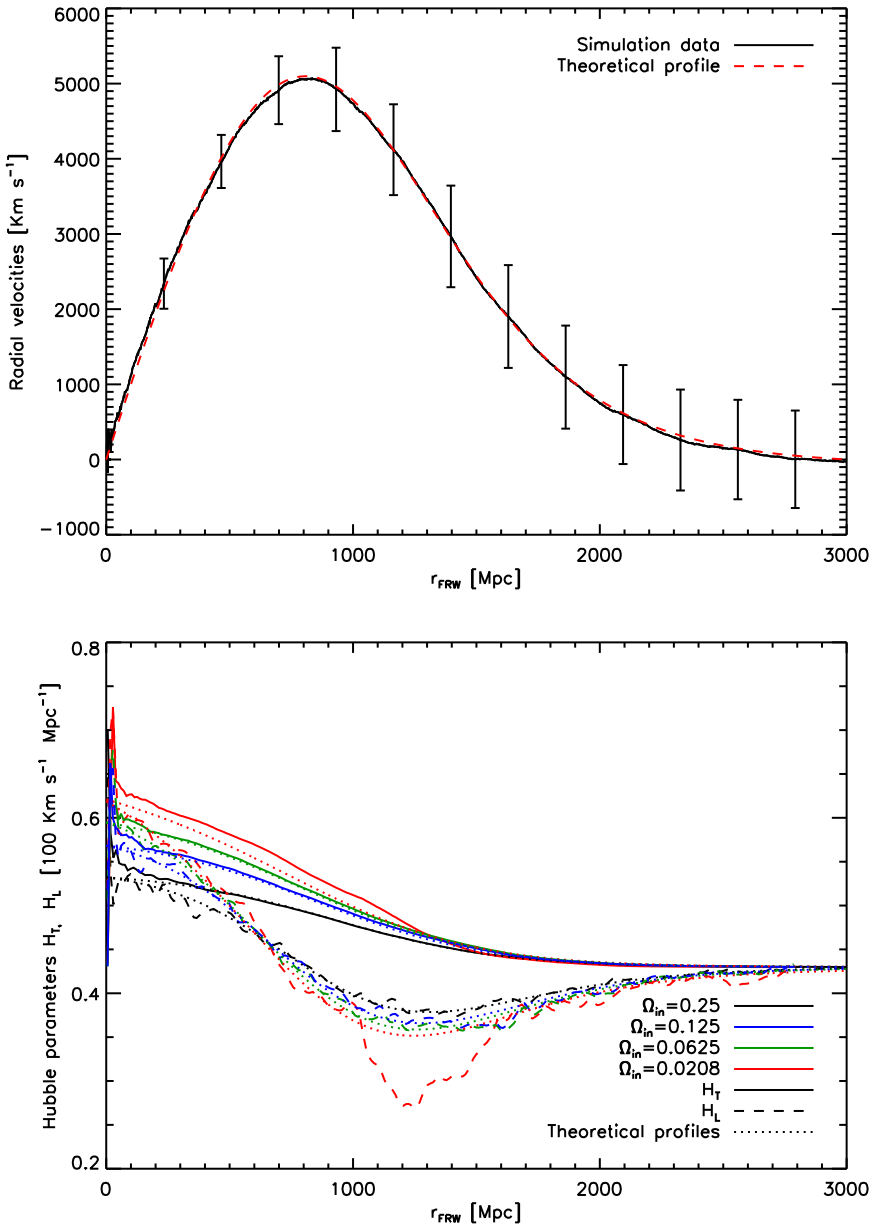

Figure 4.9: The velocity profile for the simulation $\mathcal{H}$ (top panel) and the H_T and H_L profiles for different values of $\Omega_{\rm in}$ (lower panel). In both cases, the theoretical profiles are shown with dotted lines. All curves are plotted at redshift $z = 0$.

velocities (calculated as $a^{-1}[\dot{d}_p - H_\infty d_p]$) match the data from $\mathcal{H}$. Since in the models we have simulated $(1-k(r))^{-1/2} \simeq 1$, eq. (4.48) holds and $\dot{d}_p/d_p \simeq H_T$. Using this approximation in the lower panel of fig. 4.9 we compare a range of models to the theory. Again, the difference with the theoretical curve found near $d_p = 0$ is understandable, we are shot noise dominated, and furthermore the matter perturbations displace the centre of the void slightly, while at the same time we have a formal singularity at $r = 0$ when calculating $\langle v_r \rangle / r$. From H_T we can extract H_L straightforwardly as:

$$H_L = \frac{\dot{A}'}{A'} = H_T + \frac{A}{A'} H_T', \tag{4.50}$$

which is just $H_L = H_T + r\, H_T'$ at $z = 0$, using $A_0(r) = r$. We stress though, that at $z = 0$ this is a derived parameter, and not independent of H_T. We find that all but the emptiest model $\mathcal{S}\Omega021$ match well the theoretical predictions. For $\mathcal{S}\Omega021$ the velocity is consistently higher (and the density lower) inside the void compared to theoretical predictions, and a density spike is building up near the edge of the void. This could be due to the much lower density, but it may also be a consequence of the very high starting redshift ($z_{\rm start} = 199$), that was necessary to keep the perturbations linear and the particle displacements acceptable in the initial condition.

4.3.4 Density contrast evolution

As a first attempt at using N-body simulations to study structure formation in LTB universes, we have tried to calculate the evolution of the density contrast as a function of redshift, $\delta(z) = \langle (\rho(z) - \bar{\rho})/\bar{\rho} \rangle$.

From the simulations we estimate the density contrast in a given spherical shell by calculating the r.m.s. of $\rho(r, \theta, \varphi, t)/\bar{\rho}(r, t) - 1$ over all the grid points belonging to that shell. The errors in the determination of δ were calculated as the standard deviation of the values of δ calculated in the 8 octants of each shell. The results for the simulation $\mathcal{S}49$ can be seen in fig. 4.10, where we compare the density contrast, calculated at a fixed comoving distance $r_{\rm FRW} = (1+z)\, d_p$, as a function of time (expressed in terms of redshift), with the predicted one within the simplified linear perturbation theory in LTB described in section (4.1.6) in the no-shear approximation (eq. (4.35). We also include, for comparison, the overdensity growth for an open universe, with $\Omega_{\rm M} = 0.25$, and a ΛCDM, with $\Omega_{\rm M} = 0.25$ and $\Omega_\Lambda = 0.75$. It is interesting to note that the data agree well, within error bars, with both the theoretical prediction in the LTB model and in the concordance ΛCDM, while they differ significantly from an open universe with the same matter density.

In order to better understand these differences, we have also studied the evolution of the density contrast at several distances from the centre of the void. The results are shown in fig. 4.11. Two clearly different zones can be distinguished: while the growth is proportional to $(1+z)^{-1}$ for large comoving distance, which is the expected behaviour for an EdS universe, it is significantly slower for small distances, as would occur in an open FRW universe.

4.3.5 Discussion

We have studied for the first time non-linear structure formation in large-void LTB models within an asymptotic EdS universe. By initiating large N-body simulations at high redshifts, we have been able to follow the non-linear gravitational collapse of matter structures in the presence of an underdense void that starts with a density contrast of order $\delta_m \sim 10^{-3}$ at photon decoupling (where the matter perturbations have $\delta_m \sim 10^{-5}$). We find that using

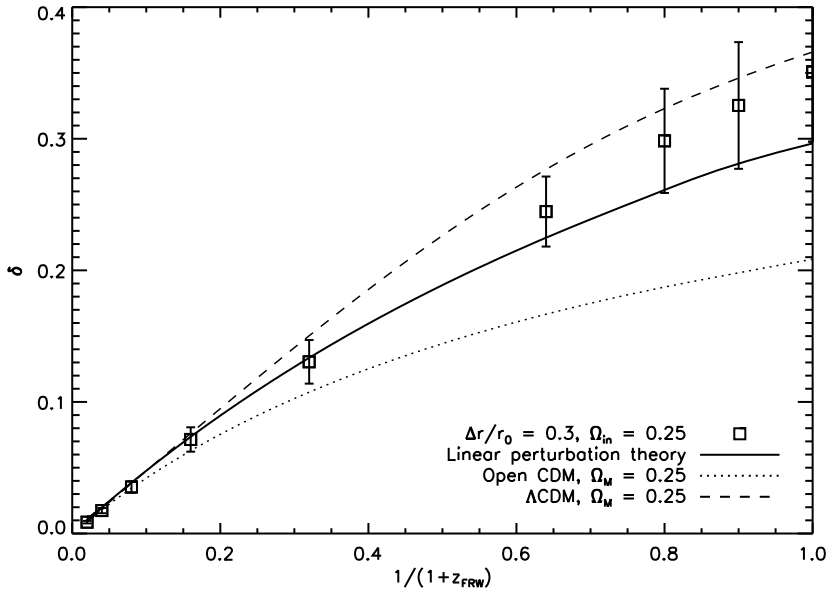

Figure 4.10: Density contrast evolution inside the void at comoving distance $r_{\rm FRW} = (1 + z)\, d_p = 280$ Mpc, for $\mathcal{S}49$, in comparison with the theoretical prediction from perturbation theory (full line). We also compare the LTB growth of density perturbations with that of Open CDM (dotted line) and ΛCDM (dashed line). The theoretical curves were normalized to have the same slope asymptotically in the past, as $a_{\rm FRW} \to 0$. Note that, even though the horizontal axis reads $1/(1 + z_{\rm FRW})$, this $z_{\rm FRW}$ only determines the cosmic time t, since the density contrast was calculated at a fixed comoving distance, and not in the lightcone.

a standard N-body code, the non-linear growth of the void underdensity follows the exact analytical solution of Einstein's equations, even for very deep voids with $\Omega_M = 0.06$ at the centre, and thus with density contrasts of order 1 with respect to the asymptotic EdS universe. Moreover, the transverse and longitudinal rates of expansion agree with the theoretical expectations, giving us confidence that the simulations are tracing the full non-linear gravitational collapse in this non-perturbative LTB background. This is furthermore evidence that N-body codes give a credible and precise description of the standard ΛCDM model, where the voids are much smaller in size, and no general relativistic corrections are needed to describe the large scale evolution.

We have also studied the evolution of the matter density contrast in such a non-trivial background, and found an analytical solution to the approximate equations for the growth of perturbations in the limit of negligible background shear which is in good agreement with the simulated results. Moreover, the comparison with OCDM and ΛCDM shows that the density contrast growth for our LTB models is very close within errors to that of the concordance ΛCDM model suggested by WMAP-7yr [32]. From our non-linear LTB N-body simulations we can potentially extract predictions for observations of large scale structure, via the two-point angular correlation function, the angular power spectrum and the growth of structure, however much work is still needed (e.g.: improving the mass resolution and including particle hydrodynamics) before any of this can be done.

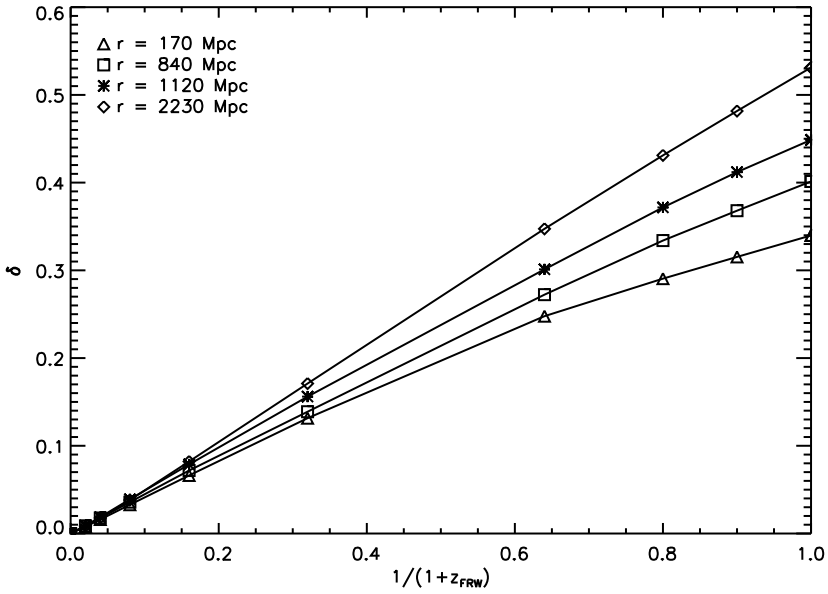

Figure 4.11: Density contrast evolution at different fixed comoving distances for $\mathcal{S}49$ simulation, as a function of the FRW scale factor, like in fig. 4.10. It is easy to distinguish between the contrast growth in a background with $\Omega \sim 1$ at large distances and with a lower Ω near the void centre.

4.4 Halo statistics in LTB models

The original motivation behind devising the method described in the previous section to perform simulations of LTB models was to be able to make safe predictions about structure formation in these models, free of large systematic uncertainties associated with the theoretical modelling. One important observable that has not yet been used to constrain LTB models is cluster number counts. The standard theoretical modelling of this quantity relies heavily on the calculation of the halo mass function: the number density of virialized dark matter clumps or "haloes". Studying the statistics of the halo population in a cosmological model can be extremely informative for structure formation, and is therefore one of the most straightforward areas of applicability of our simulations. This study motivated the publication [8] and is the topic of this section.

This analysis could be interesting not only for LTB purposes. It may also give some hints about structure formation in astrophysical voids, and about environmental effects on the properties of haloes. In this sense, the advantage of our simulations is the fact that we have absolute control over the properties and evolution of the void.

As has been said in the previous section, the mass resolution of our simulations is not very good. However, in the case of $\mathcal{H}$ it is enough to obtain information about halo abundances, although not sufficiently small to study the internal properties of haloes, due to the small number of particles that form each of them.

In section 4.4.1 we describe our theoretical model for the halo mass function in LTB models. We describe the algorithm used to find and characterize the haloes in section 4.4.2. In sections 4.4.3 and 4.4.4 we show our results concerning halo abundances and orientations respectively. Finally we discuss these results in section 4.4.5.

4.4.1 The mass function in LTB models

Most of the information about the non-linear accretion of dark matter haloes is encoded in the mass function $n(M)\,dM$: the comoving number density of haloes with mass $M \in (M, M+dM)$. The first theoretical description of the mass function was developed by Press and Schechter [104] (PS hereon) and later re-derived and extended by Bond et al. in the so-called excursion set formalism [111]. Within this framework the abundance of haloes can be predicted as the abundance of points in space in which the linear density contrast δ smoothed over a scale corresponding to the mass M has crossed the spherical collapse threshold $\delta_c = 1.686$. Although the PS prediction describes qualitatively well the mass function, it fails to reproduce its details (overpredicting the density of low mass objects and underpredicting massive ones). Nevertheless it is a remarkable achievement that one can estimate the abundance of non-linear structures using only linear perturbation theory and the assumption that δ is Gaussianly distributed. The PS formula has been perfected using ellipsoidal collapse and empirical parametrizations [227, 109], so that $n(M)$ can be calculated to very good accuracy, often using one of the main results from this formalism: the mass function should be a universal (cosmology-independent) function of the variance of the linear density contrast field $\sigma(M,z)$ [107]. Here we will use

$$n(M,z) = \frac{\rho_{\rm M}}{M} g(\sigma) \left| \frac{d\,\ln\,\sigma}{dM} \right|, \tag{4.51}$$

where $\sigma \equiv \sigma(M,z)$ and $g(\sigma)$ is given by [228]:

$$g(\sigma) \equiv \frac{a\,b\,\nu^b + 2\,c\,\nu^2(1+a\,\nu^b)}{(1+a\,\nu^b)^2} \exp(-c\,\nu^2),$$
$$(a,b,c) = (1.529, 0.704, 0.412), \quad \nu \equiv \delta_c/\sigma.$$

We have also tried other parametrizations of the mass function [227, 109] and checked that our results did not depend significantly on this choice.

We follow the same rationale in order to calculate the mass function of haloes at a given r and t in an LTB model: since the simulated void arises from a purely growing mode (i.e.: the Big Bang time is homogeneous), and perturbations grow in a self-similar fashion, it is reasonable to assume that, in order to calculate the variance of $\delta_{\rm M}$ at (t,r), we should rescale the variance $\sigma_{\rm out}(M,z)$ of the density perturbations outside the void, by a factor

$$f(t,r) = \frac{\delta_\alpha(t,r)}{\delta_\alpha(t, r\to\infty)}\,, \tag{4.52}$$

where the density contrast is computed theoretically according to eq. (4.37), and evaluated in fig. 4.12, with and without the shear correction. Thus, our model for the mass function $n(M,z,r)$ at a given radius r is eq. (4.51) with $\sigma(M,z)$ substituted by $\sigma_{\rm out}(M,z)\,f(t,r)$.

Note however, that our main results will be quoted in terms of the cumulative mass function within a sphere of radius r centred at the origin of the LTB patch:

$$n(>M, <r, z) \equiv \frac{3}{4\pi\,r^3} \int_0^r r'^2 dr' \int_M^\infty dM'\, n(M', r', z), \tag{4.53}$$

since this observable has better statistics.

4.4.2 The halo catalogue

Our aim is to extract, from the dark matter particles of the simulation, the virialized structures or haloes. There exist different halo-finding algorithms in the literature, and their accuracies at describing different halo properties have been studied in detail in [229].

A very popular halo-finding method is the so-called Friends-of-Friends (FoF) algorithm. In this case haloes are formed by particles which are separated by less than a given distance (the "linking length") from other particles of the halo. Generally the linking length is defined as a fraction of the mean inter-particle distance $d_{\rm min} = b\,\bar{d}$. A common choice for the linking length parameter is $b = 0.2$ [230].

Another popular class of halo finders are the so-called "spherical overdensity" finders. In this case the density field is computed from the particle content and the positions of possible halo centres are located as local overdensities. Then the gravitationally bound particles around these centres are extracted, and the extent and mass of each halo is computed as:

$$M(r_{\rm vir}) = \frac{4\pi}{3}\, r_{\rm vir}^3\, \rho_c\, \Delta, \tag{4.54}$$

where Δ is the overdensity parameter. Even though the value for Δ predicted by the top-hat spherical collapse theory is $\Delta_{\rm th} = 178$, it is common to use $\Delta = 200$ or $\Delta = 500$ (which lead to different definitions of the halo mass).

The halo catalogue used in the present analysis has been extracted using the AMIGA halo finder `AHF` [231]. `AHF` is a spherical overdensity (SO) halo finder that maps the particle content to an adaptively smoothed density field. For our catalogue we have used an overdensity parameter $\Delta = 200$.

Each halo is modelled as an ellipsoid and its axis lengths (a, b, c) and orientations ($\hat{\mathbf{e}}_a$, $\hat{\mathbf{e}}_b$, $\hat{\mathbf{e}}_c$) are found as the eigenvalues and (normalized) eigenvectors of the inertia tensor

$$I_{ij} \equiv \sum_p x_p^i\, x_p^j, \tag{4.55}$$

where p runs over all particles in the halo.

`AHF` was run on our highest resolution simulation $\mathcal{H}$, locating about 10^6 haloes containing 20 simulation particles or more. All these have been used to study the mass function, but only those with more than 50 particles were used to analyze their orientations.

4.4.3 Halo abundances

Once the haloes have been identified, we can compute $n(>M, <r, z)$ merely by counting the number of haloes with mass above M inside a sphere of radius r and dividing by the comoving volume of this sphere. The results are shown in figure 4.12. It is easy to see (top panel) that, while approximating the growth of perturbations by its 0-th order in ϵ (eq. (4.35)) yields a reasonably good fit at small and large radii (where the background shear vanishes), it fails to reproduce the halo abundances at intermediate radii. In the bottom panel we can see, however, that adding a non-zero 1-st order correction (eq. (4.37)) solves this problem. Furthermore, we have found that this correction seems to be almost independent of r and t, with $\alpha \sim 2$.

We have quantified the goodness of fit of our approach (with and without the shear correction term) using the measure:

$$\xi_N^2(r, z) \equiv \frac{1}{N-1} \sum_i^N \left(\frac{n(>M_i, <r, z) - n_i(r)}{n(>M_i, <r, z)} \right)^2. \tag{4.56}$$

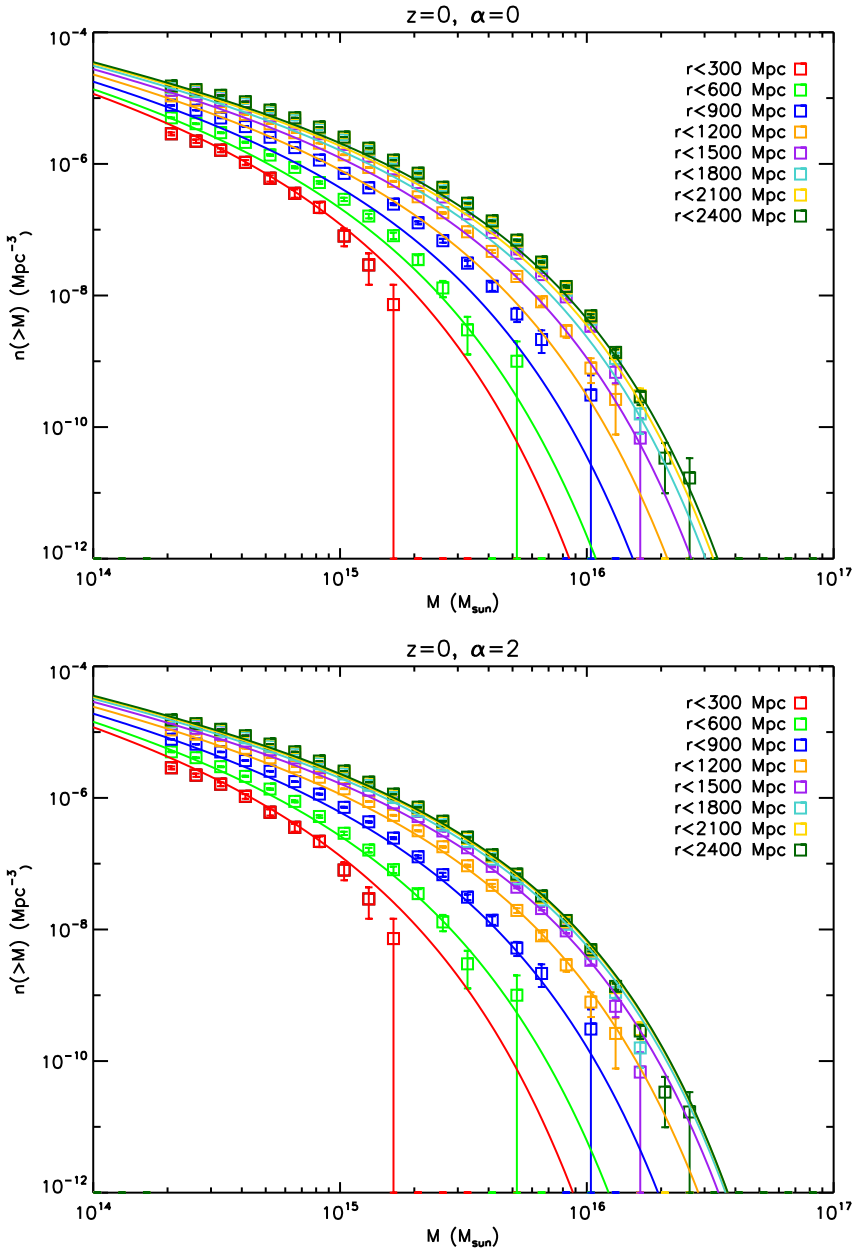

Figure 4.12: Cumulative mass function for haloes inside spheres of different radii compared with the theoretical prediction without the background shear correction ($\alpha = 0$, top panel) and with a first-order shear correction, where $\alpha = 2 \forall\, r,\, t$ (bottom panel).

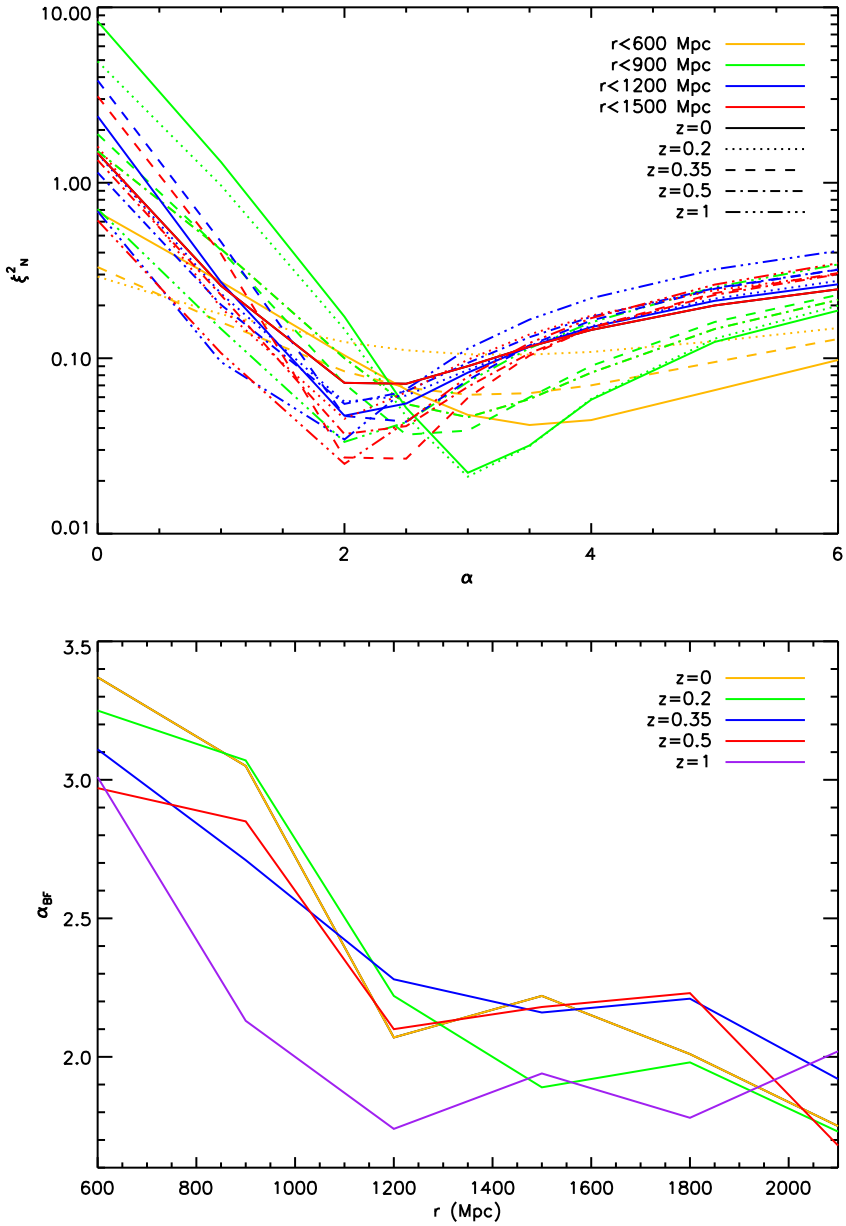

Figure 4.13: Top panel: dependence of ξ^2_N (eq. (4.56)) with the value of the shear correction term α at different r and t. Bottom panel: best fit value of α for different radii and redshifts, showing a mild evolution with r.

r (Mpc)	$\xi_N^2(\alpha=0)$	$\xi_N^2(\alpha=2)$
300	0.093	0.112
600	0.682	0.104
900	8.321	0.173
1200	2.394	0.047
1500	1.474	0.072
1800	0.478	0.050
2100	0.177	0.032
2400	0.095	0.061

Table 4.2: Goodness of fit of our theoretical approach with and without a first-order shear correction for different radii at $z = 0$. Similar results hold at all other redshifts (see table 4.3).

z	$\xi_N^2(\alpha=0)$	$\xi_N^2(\alpha=2)$
0	1.620	0.072
0.2	1.163	0.065
0.35	1.411	0.042
0.5	0.667	0.047
1	3.494	0.824

Table 4.3: Goodness of fit of our theoretical approach with and without a first-order shear correction for different redshifts (summing over all radii).

Here $n(>M,<r,z)$ is given in section 4.4.1, $n_i(r)$ is the cumulative mass function obtained from the simulation for a mass M_i within a sphere of radius r, and N is the number of mass bins. Figure 4.13 (top panel) shows the value of ξ_N^2 for different choices of α at different r and z. A value of $\alpha \sim 2$, found as the median of the best-fit values for all the calculated curves, gives a good fit in all cases with only a very mild dependence on r and t (shown at the bottom panel of fig. 4.13). The improvement due to the shear-correction term can also be seen in table 4.2, in which we have calculated the goodness of fit with and without the shear correction for different radii at $z = 0$. Table 4.3 shows the same result for different redshifts summing over all radii. This improvement is especially evident at intermediate radii, where ϵ is larger and therefore its effects are more important.

It would be extremely interesting to investigate whether and how the value of α depends on the cosmological parameters: if this parameter turned out to be independent of the void model, one should be able to predict its value from some approximation in perturbation theory. However, this is work in progress and we defer the presentation of it to future work that will also make use of better resolved simulations.

4.4.4 Halo orientations

In addition to the mass spectrum of haloes and the proposed modification to its theoretical prediction based upon the PS formalism, we would additionally like to present a study of another property of dark matter haloes found in our simulation, namely the orientation of the halo with respect to an observer placed at the centre of the void. To this extent, the radial dependence of this quantity has been calculated by binning the haloes in spherical shells thick enough to contain at least 1000 haloes.

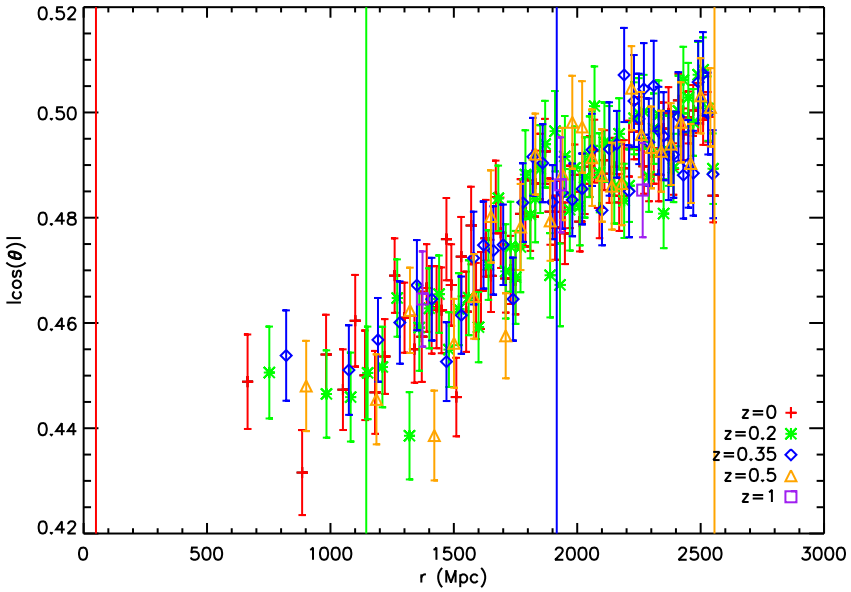

Figure 4.14: Radial dependence of halo orientations. The vertical lines show the distance corresponding to each redshift to facilitate the interpretation of the lightcone signal.

For this study we have restricted our halo catalogue to those with at least 50 simulation particles, since we have found that only then can we trust the estimation of the relative axes with sufficient accuracy. This drastically reduces our statistics at low radii, basically losing all data below 500 Mpc.

We quantify the orientation of the haloes in terms of the cosine of the angle between the halo's main axis $\hat{\mathbf{e}}_a$ and the line of sight of the central observer $\hat{\mathbf{e}}_r$. For a homogeneous background, halo orientations should be randomly distributed, in which case the expected value of the random variable $|\cos(\theta)| \equiv |\hat{\mathbf{e}}_a \cdot \hat{\mathbf{e}}_r|$ would be $1/2$. We can see in figure 4.14 that, indeed, we recover that result outside the void, in the FRW background. However, close to the void wall ($\sim$ 1100 Mpc) there seems to be a slight preference for haloes to be oriented in the plane transverse to the line of sight ($\langle|\cos(\theta)|\rangle < 0.5$). We can also see that this signature seems to vary very little with time within the range of redshifts we have studied, and therefore could be observed in the lightcone.

We believe this behaviour can be understood (at least qualitatively) again as an effect of the background shear: the fact that the expansion rate is larger in the transverse direction than in the longitudinal direction. The expansion rate enters the perturbation equations as a damping term for the peculiar velocities, and the fact that $H_L < H_T$ makes this damping larger for longitudinal modes. Namely:

$$\partial_T \dot{\Psi} + (H_T - \Delta H)\,\partial_T \Psi \propto \rho\, v_T, \qquad \partial_L \dot{\Psi} + (H_T + \Delta H)\,\partial_L \Psi \propto \rho\, v_L, \tag{4.57}$$

where $\Delta H = H_T - H_L$. Thus transverse infall velocities will be, in principle, larger, and haloes will be preferably oriented in this direction. This effect is more relevant close to the void wall, where ΔH reaches a maximum, and will disappear for $r \ll r_0$ and $r \gg r_0$. We note that the same result has been found in [232], although in astrophysical voids of much smaller size.

4.4.5 Discussion

We have extracted the halo content from an LTB N-body simulation and analyzed the halo abundances at different masses and radii, as well as their orientations with respect to the line of sight. The main conclusions from this study are:

- The theoretical description of the halo mass function in FRW cosmologies can be fully extended to LTB void models by adding just one parameter (α) that accounts for the effect of the background shear on the evolution of matter density perturbations.
- The value of this parameter (~ 2) seems to be constant in time and only mildly dependent on the position in the void. Whether this value depends weakly on the void model is still work in progress.
- Haloes close to the void wall are preferably oriented in the transverse plane. This can be understood as an effect of the different longitudinal and transverse expansion rates.

If the shear correction turns out to be practically independent of the void model parameters, halo abundances could potentially be used to constrain the amount of background shear, a crucial test for general inhomogeneous cosmological models. A toy model that has been considered in the past in connection with LTB scenarios, and which could benefit from our analysis of background shear, is the Swiss-cheese model [233, 234, 235]. We leave for the future such investigation. It is also worthwhile exploring whether our results apply to voids of astrophysical scales, of tens of Mpc, since they could have an effect on the modelling of the environmental dependence of dark matter halo properties, as well as the backreaction of non-linear gravitational collapse on the background evolution.

CHAPTER 5

Conclusions

In this thesis we have explored two different topics, related to the late-time evolution of the Universe and the problem of dark energy: the measurement of the sound horizon scale in galaxy surveys and the formation of large-scale structure in LTB void models.

Measuring the BAO scale from galaxy redshift surveys. The measurement of the BAO scale is an extremely robust probe of the distance-redshift relation, and therefore a very useful tool to study the nature of dark energy. Two distinct observables can be measured in BAO studies: the angular scale of the sound horizon, which gives information about the angular diameter distance, and the radial BAO scale, which is directly related to the expansion rate. In spite of this, the most relevant BAO detections have measured an angle-averaged combination of these two scales. While this approach makes a very efficient use of the data, since all pairs of galaxies are taken into account, regardless of their orientation, forthcoming galaxy surveys such as DES-spec, Euclid or BigBOSS should obtain enough statistics to measure both quantities separately. Another drawback of the standard approach to BAO is the fact that it has usually been measured from the three-dimensional two-point correlation function, a quantity for which the 3D distances between galaxies must be used. This means that the galaxy redshifts must be translated into distances using a fiducial cosmological model, and it is not clear that by doing this we will not bias the final cosmological constraints. It is true that the effects due to assuming an incorrect cosmology that is close to the true one can be understood in terms of the Alcock-Paczynski effect, however there is no clear measure of what we mean by "close", and there is no conclusive study in the literature as to the effects of this assumption on the final BAO results. Most importantly, even though we have some indications as to what the "true" cosmology is, most of these indications come either from low redshift observables such as SNe or galaxy clustering, or from early-Universe phenomena, such as the CMB anisotropies. We must be very careful before we assume a particular fiducial cosmology when analyzing the future data from galaxy surveys, Lyman-α experiments and 21cm observations, which will cover the unexplored intermediate redshifts.

For these reasons we have proposed a method to measure the angular and radial BAO scales independently from galaxy redshift surveys. The method uses purely observable quantities (redshifts and angles) and its results are therefore unbiased towards any model. Our analysis is based on an empirical parametrization of the projected angular and radial correlation functions, which fits the tilt of the 2PCF with a smooth function and uses a Gaussian term to determine the position of the BAO.

The projection of the galaxy distribution on thick redshift shells in order to measure the angular two-point correlation function, shifts and broadens the angular BAO peak. However, we have shown that these projection effects can be incorporated in the method. This method can also be used to measure the angular scale of the sound horizon from photometric galaxy surveys. However, the use of photometric redshifts introduces a dominant source of systematic uncertainty.

In the case of the radial correlation, projection effects are limited to small scales, and do not affect the BAO scale. The detection of the radial BAO feature is however limited by the survey volume, since cosmic variance plays a dominant role. This is not surprising, since only a small fraction of galaxy pairs are oriented close to the line of sight.

All in all, the method is extremely insensitive to systematic uncertainties (except for the photometric case), and provides a very robust determination of the BAO scale. We have tested the performance of our method using a simulated galaxy catalog provided by the MICE collaboration, and we have verified that very significant cosmological constraints can be obtained from BAO data alone when radial and angular measurements are combined. Furthermore, we have compared these constraints with the ones obtained through the standard BAO analysis. Both methods seem to yield comparable results, with the added advantage that the proposed method does not assume any fiducial cosmology, a fact that may be extremely important when constraining alternative models of dark energy.

Large-scale structure formation in LTB models. LTB void models have been proposed as an alternative explanation to the apparent accelerated expansion of the Universe. In these models the observer is located close to the centre of a spherical, Giga-parsec sized void, and the apparent acceleration in the rate of expansion is only a result of the misinterpretation of the observational data in terms of a homogeneous cosmology. Even though these models may be outrageous in their explicit violation of the CP, trying to rule them out using only observational data is a very healthy scientific exercise. This endeavour has created a lot of scientific activity in the last years, and most LTB models can now be considered to be ruled out.

Our contribution to the study of void models has been to devise a technique to perform N-body simulations of them. This would enable us to better understand the formation of structure in these models, a topic that is highly non-trivial even in the linear regime, given that perturbation theory in an LTB background is far more complicated than the standard case. The technique used to simulate LTB models was to understand the void as arising from a perturbation at early times of very large scale but small in amplitude. This perturbation evolves via gravitational collapse into the highly non-linear void needed at late times. Since no new physics is needed (the only matter component is dust and gravity is ordinary GR), we only need to generate the initial conditions for the simulation correctly, and the gravitational evolution will take care of the rest. Since the exact solutions for the LTB metric are known due to its spherical symmetry, this task could easily be done by modifying an existing initial condition generator to take into account both the small-scale perturbations due to the primordial fluctuations and the large-scale void. Care must be taken when using this technique: because cosmological N-body simulations have periodic boundary conditions, our method would not simulate just one LTB void, but a network of voids consisting of the mirror images of the intended one. In order to make the interaction between the different voids negligible, the box size must be considerably larger than the size of the void, and we verified that by doing so all boundary effects disappear. This exposes one of the caveats of the method: since realistic voids are huge ($\mathcal{O}(2-3)\,\mathrm{Gpc}$), the box must have a considerable size, and therefore a large number of particles is necessary in order to resolve small structures.

Thus, depending on the range of scales that need to be explored, simulating an LTB void may be computationally costly. We verified that our simulations were able to follow the evolution of the large-scale void correctly according to the exact solutions of the LTB metric, both for the density and velocity fields, and for a wide range of LTB models.

One of the key characteristics of LTB models is that, at the background level, the evolution at a fixed comoving distance from the centre of the void is equivalent to the evolution of an FRW universe with some effective cosmological parameters. This result however does not hold at the perturbative level. LTB models have a non-zero background shear component: due to the fact that these models have two different expansion rates, along the line of sight and in the transverse direction, extended objects expand in size, but also change in shape. This shear affects the way in which perturbations grow and causes departures from the local effective FRW solution. Our simulations were used to address this effect. First, we studied the growth of perturbations through the variance of the density contrast field at different radii, and verified that, qualitatively, perturbations grow as they would in an open universe inside the void, while their growth is compatible with that of an EdS cosmology outside it. This result was addressed quantitatively by studying the statistics of the dark matter halos in these simulations. The halos were extracted using a spherical-overdensity halo finder and the abundance of halos of different mass was studied for spheres of different radii centered at the void. We found that the halo abundances were not compatible with the growth of structure predicted by a local FRW universe at every radii, and that a correction proportional to the amount of background shear is necessary. This correction seems to be of the same order for all radii in the simulation for which the analysis was done. If independent of the parameters of the background LTB model, then halo abundances could be used to trace the amount of background shear, a clear indication of an inhomogeneous expansion.

Final remark. Through my Ph.D. studentship I have realized that an open mind is necessary to address the scientific enigmas of cosmology. In the case of dark energy, even though the standard cosmological model seems to obtain the best agreement with the data, with the smallest number of free parameters, alternative models must be taken into account and analyzed, since they may contain the key to understanding the shortcomings of the standard theory. These models should be explored and judged only in the light of observational evidence, which must be free of assumptions, preconceptions and prejudices. The abundance of past and present data, and the promise of new observations to come in the near future give us hope that this rigorous analysis can be done, and that, in doing so, many of the present open questions in cosmology will be closed, while many more will be opened.

APPENDIX A

Statistics of the matter distribution

Throughout this thesis we have dealt with perturbations on different background quantities, such as the overdensity field δ, the peculiar velocity $\mathbf{v}$ or the Newtonian potential Φ. As we understand them, these fields are but one realization of an underlying statistical ensemble, the origin of which can be understood, for example, through the theory of inflation (see [236, 67]). The properties of these fields can then be described in a statistical sense by studying their one-point probability distribution and the correlations between the fields at different points. In section A.1 we will define some of the key concepts in this description, and state some important results about the main statistical quantities of random fields.

In observational cosmology, often discrete tracers of the underlying continuous fields are used to study them, the most evident example being the galaxy distribution. Section A.2 deals with the description and main properties of these point processes.

A.1 Random fields

For the sake of generality we will speak here of a general random field $f(\mathbf{x})$ with zero mean ($\langle f(\mathbf{x})\rangle = 0$), but these results can be applied to any cosmological fluctuation.

A.1.1 Definitions and important results

Random field. A random field is an $\mathbf{x}$-dependent random variable

$$(f(\mathbf{x}), p[f(\mathbf{x})])\,, \tag{A.1}$$

where $p[f(\mathbf{x})]$ is the related probability density functional (p.d.fn.). It is often useful to work with the Fourier transform of the field, which we define as

$$f_{\mathbf{k}} \equiv \int \frac{dx^3}{(2\pi)^{3/2}} f(\mathbf{x})\, e^{-i\,\mathbf{k}\cdot\mathbf{x}}. \tag{A.2}$$

In general the Fourier transform will be a complex field, but the condition

$$f_{\mathbf{k}} = f^*_{-\mathbf{k}} \tag{A.3}$$

must be satisfied for $f(\mathbf{x})$ to be real.

The probability that the field f will adopt a given configuration between $f(\mathbf{x})$ and $f(\mathbf{x}) + df(\mathbf{x})\ \forall\,\mathbf{x}$ is given by

$$d\mathcal{P} = p[f(\mathbf{x})] \prod_{\mathbf{x}} df(\mathbf{x}) \equiv p[f(\mathbf{x})]\,\mathcal{D}f. \tag{A.4}$$

Homogeneous field. A random field $f(\mathbf{x})$ is said to be homogeneous iff any correlation function depends only on the differences between arguments:

$$f \text{ homogeneous } \iff \forall\,\mathbf{z},\ \langle \prod_i f(\mathbf{x}_i)\rangle = \langle \prod_i f(\mathbf{x}_i + \mathbf{z})\rangle \tag{A.5}$$

(i.e.: correlators are invariant under translations).

In the same way we define isotropic fields as those whose correlators are invariant under $SO(3)$ transformations of the field arguments. For example, it is easy to show that, for a homogeneous and isotropic field, the two-point correlator depends only on the distance:

$$\xi(\mathbf{x},\mathbf{y}) \equiv \langle f(\mathbf{x}) f(\mathbf{y})\rangle = \xi(|\mathbf{x}-\mathbf{y}|) \tag{A.6}$$

Uncorrelated field. A random field is uncorrelated at large distances iff

$$\begin{aligned}\forall\, p,q\ ,\ \lim_{u\to\infty} \langle \prod_{i=1}^{p} f(\mathbf{x}_i + \mathbf{u}) \prod_{j=1}^{q} f(\mathbf{y}_j - \mathbf{u})\rangle &= \langle \prod_{i=1}^{p} f(\mathbf{x}_i + \mathbf{u})\rangle \langle \prod_{j=1}^{q} f(\mathbf{y}_j - \mathbf{u})\rangle \\ &= \langle \prod_{i=1}^{p} f(\mathbf{x}_i)\rangle \langle \prod_{j=1}^{q} f(\mathbf{y}_j)\rangle,\end{aligned} \tag{A.7}$$

where we have assumed homogeneity in the last line.

Gaussian fields. A random field $f(\mathbf{x})$ is said to be Gaussian iff any correlation function is equal to the sum over pairings of the product of the averages of each pair (and 0 if there is an odd number of fields). This definition is much clearer with an example:

$$\begin{aligned}\langle f(\mathbf{x}_1) f(\mathbf{x}_2) f(\mathbf{x}_3) f(\mathbf{x}_4)\rangle = &\langle f(\mathbf{x}_1) f(\mathbf{x}_2)\rangle \langle f(\mathbf{x}_3) f(\mathbf{x}_4)\rangle + \\ &\langle f(\mathbf{x}_1) f(\mathbf{x}_3)\rangle \langle f(\mathbf{x}_2) f(\mathbf{x}_4)\rangle + \langle f(\mathbf{x}_1) f(\mathbf{x}_4)\rangle \langle f(\mathbf{x}_2) f(\mathbf{x}_3)\rangle.\end{aligned} \tag{A.8}$$

Since the average of the product of an odd number of f's is zero, the two-point function

$$\xi(\mathbf{x},\mathbf{y}) \equiv \langle f(\mathbf{x}) f(\mathbf{y})\rangle \tag{A.9}$$

alone determines all the moments of the probability distribution for a Gaussian field. An important consequence of this property is that the joint probability that the field takes some given values at n different points is given by a multivariate Gaussian:

$$p(\overrightarrow{f}(\mathbf{x}) = \overrightarrow{f}_0) \prod_{i=1}^{n} df_i = \frac{1}{\sqrt{(2\pi)^n \det(\hat{C})}} \exp\left[-\frac{1}{2}\overrightarrow{f}_0^{\,T} \cdot \hat{C}^{-1} \overrightarrow{f}_0\right] \tag{A.10}$$

where $\overrightarrow{f}(\mathbf{x}) = (f(\mathbf{x}_1), ..., f(\mathbf{x}_1))$ and $\hat{C}$ is the covariance matrix

$$C_{ij} = \langle f(\mathbf{x}_i) f(\mathbf{x}_j)\rangle \tag{A.11}$$

The ergodic theorem. Let $f(\mathbf{x})$ be a homogeneous random field uncorrelated at large distances. Qualitatively, the ergodic theorem tells us that ensemble averages and spatial averages are equivalent as long as the region of space over which we average is large enough. Quantitatively, the ergodic theorem can be stated as follows: let $W(z)$ be a Gaussian window function of width R (it does not have to be necessarily Gaussian, any window function with the same characteristics would do)

$$W(\mathbf{z}-\mathbf{z}_0) \equiv (2\pi R^2)^{-3/2} \exp\left(-\frac{(\mathbf{z}-\mathbf{z}_0)^2}{2R^2}\right), \tag{A.12}$$

and let us define spatial averaging as

$$\langle \prod_i f(\mathbf{x}_i) \rangle_S^R \equiv \int dz^3 W(\mathbf{z}) \prod_i f(\mathbf{x}_i + \mathbf{z}). \tag{A.13}$$

We define the error function

$$\Delta_R^2 \equiv \left\langle \left(\langle \prod_i f(\mathbf{x}_i) \rangle_S^R - \langle \prod_i f(\mathbf{x}_i) \rangle \right)^2 \right\rangle. \tag{A.14}$$

The ergodic theorem is

$$\lim_{R\to\infty} \Delta_R = \mathcal{O}(R^{-3/2}). \tag{A.15}$$

The proof of this can be found in [236].

A.1.2 Filtering

In some cases we may be interested in studying the statistics of a random field within a given range of scales, and we would like to be able to filter away all scales smaller than a given threshold R. This can be done by convolving the field with a window function whose width corresponds to this minimum scale

$$f_R(\mathbf{x}) \equiv \int dy^3 \, f(\mathbf{x}+\mathbf{y}) \, \tilde{W}(\mathbf{y}; R). \tag{A.16}$$

Two popular types of filters are "Top-Hat" filters and "Gaussian" filters:

$$\tilde{W}_{\rm TH}(x;R) \equiv \frac{3}{4\pi\, R^3}\Theta(R-x), \quad \tilde{W}_{\rm G}(x;R) \equiv \frac{1}{(2\pi\, R^2)^{3/2}} \exp\left(-\frac{x^2}{2\, R^2}\right). \tag{A.17}$$

Using the convolution theorem, this operation is simplified in Fourier space, and $f_{R,\mathbf{k}} = W(kR)\, f_{\mathbf{k}}$ ($W(kR)$ is the Fourier transform of $\tilde{W}(x,R)$). The variance of the smoothed field can be related to the power spectrum of the unfiltered field (see next section) via

$$\sigma_R^2 \equiv \langle |f_R(x)|^2 \rangle = \frac{1}{2\pi^2} \int_0^\infty k^2 \, |W(kR)|^2 \, P(k) \, dk. \tag{A.18}$$

For convenience, the Fourier transforms of the two window functions defined above are

$$W_{\rm TH}(x) = 3\, \frac{\sin x - x\, \cos x}{x^3}, \quad W_{\rm G}(x) = e^{-x^2/2}. \tag{A.19}$$

A.1.3 The two-point correlation function and the power spectrum

The lemmas in this section will be given without proof, however most of them are straightforward or can be found in [237].

The power spectrum

Lemma: If $f(\mathbf{x})$ is a homogeneous random field, then the Fourier-space 2-point correlation function satisfies:

$$\langle f_{\mathbf{k}} f_{\mathbf{q}} \rangle \propto \delta^{\mathcal{D}}(\mathbf{k}+\mathbf{q}) \tag{A.20}$$

Definition: For a homogeneous random field the power spectrum $P_{\mathbf{k}}$ is defined by

$$\langle f_{\mathbf{k}} f_{\mathbf{q}} \rangle = \delta^{\mathcal{D}}(\mathbf{k}+\mathbf{q}) P_{\mathbf{k}} \tag{A.21}$$

Two-point correlation function

The two-point correlation function of a random field $f(\mathbf{x})$ is

$$\xi(\mathbf{x}_1, \mathbf{x}_2) \equiv \langle f(\mathbf{x}_1)\, f(\mathbf{x}_2) \rangle. \tag{A.22}$$

If the field is homogeneous, the correlation function will only depend on the difference $\mathbf{r} \equiv \mathbf{x}_1 - \mathbf{x}_2$

$$\xi(\mathbf{r}) = \langle f(\mathbf{x})\, f(\mathbf{x}+\mathbf{r}) \rangle. \tag{A.23}$$

Furthermore, for a homogeneous and isotropic field this will only depend on $r \equiv |\mathbf{r}|$.

Lemma: The power spectrum is the Fourier transform of the two-point correlation function of a homogeneous random field:

$$\begin{aligned} P_{\mathbf{k}} &= \int dr^3\, \xi(\mathbf{r}) e^{-i\mathbf{k}\cdot\mathbf{r}}, \\ \xi(\mathbf{r}) &= \frac{1}{(2\pi)^3} \int dk^3\, P_{\mathbf{k}}\, e^{i\mathbf{k}\cdot\mathbf{r}} \end{aligned} \tag{A.24}$$

For a homogeneous and isotropic random field, $\xi(\mathbf{r}) = \xi(r)$ and $P_{\mathbf{k}} = P_k$, and the formulas above can be simplified further

$$\begin{aligned} P_k &= 4\pi \int_0^\infty \xi(r)\, j_0(kr)\, r^2\, dr, \\ \xi(r) &= \frac{1}{2\pi^2} \int_0^\infty P_k\, j_0(kr)\, k^2\, dk \end{aligned} \tag{A.25}$$

where $j_0(kr) \equiv \frac{\sin(kr)}{kr}$.

A.1.4 Angular correlations functions and C_l's

Given a random field in $\mathbb{R}^3$, $f(\mathbf{x})$ (which, for a better interpretation may be identified with, say, the number density contrast $\delta(\mathbf{r})$ of some type of objects), its projection on the sky is defined as

$$f(\hat{\mathbf{n}}) \equiv \int_0^\infty f(\mathbf{r})\, \psi(r)\, r^2 dr, \quad \hat{\mathbf{n}} \equiv \frac{\mathbf{r}}{r}, \tag{A.26}$$

where the function $\psi(r)$ is the selection function, which quantifies our capability to measure f to a given comoving depth r (in the analogy above, this would be the comoving density of objects in a given survey, given by the integrated luminosity function). The selection function is normalized to unity

$$\int_0^\infty \psi(r)\, r^2 dr = 1. \tag{A.27}$$

Usually the selection function is given in terms of redshift $\phi(z)$ instead of comoving distance r. This change of variables is easy:

$$\phi(z) = \psi(r(z))\frac{r^2(z)}{H(z)} \tag{A.28}$$

As before, $f(\hat{\mathbf{n}})$ can be expanded in a function series. In the flat three-dimensional case the natural choice were plane waves $e^{i\mathbf{k}\cdot\mathbf{r}}$, but in $\mathcal{S}^2$ the natural basis are spherical harmonics $Y_{lm}(\hat{\mathbf{n}})$. Thus, for the projection of a random field $f(\hat{\mathbf{n}})$ we define the a_{lm}'s as:

$$\begin{aligned} f(\hat{\mathbf{n}}) &= \sum_{l=0}^{\infty}\sum_{m=-l}^{l} a_{lm}\, Y_{lm}(\hat{\mathbf{n}}), \\ a_{lm} &= \int d\Omega^2 f(\hat{\mathbf{n}})\, Y^*_{lm}(\hat{\mathbf{n}}) \end{aligned} \tag{A.29}$$

Since $Y^*_{lm}(\hat{\mathbf{n}}) = Y_{l-m}(\hat{\mathbf{n}})$, for $f(\hat{\mathbf{n}})$ to be real, the a_{lm}'s must satisfy $a^*_{lm} = a_{l-m}$. We see that in $\mathcal{S}^2$ the a_{lm}'s play the role of the Fourier transform in the three-dimensional case.

Review of spherical functions:

- Legendre polynomials:

$$L_l(x) = \frac{1}{2^l\, l!}\frac{d^l}{dx^l}(x^2-1)^2, \tag{A.30}$$

$$\int_{-1}^{1} dx\, L_l(x)\, L_{l'}(x) = \frac{2}{2l+1}\delta^{\mathcal{K}}_{ll'}. \tag{A.31}$$

- Associated Legendre polynomials:

$$L_{lm}(x) = (1-x^2)^{m/2}\frac{d^m}{dx^m}L_l(x), \quad (m = 0, ..., l), \tag{A.32}$$

$$L_{l-m} = \frac{(l-m)!}{(l+m)!}\, L_{lm}, \tag{A.33}$$

$$\int_{-1}^{1} L_{lm}(x)\, L_{l'm}(x)\, dx = \frac{2}{2l+1}\frac{(l+m)!}{(l-m)!}\,\delta^{\mathcal{K}}_{ll'}. \tag{A.34}$$

- Spherical harmonics (here $\hat{\mathbf{n}} \leftrightarrow (\theta, \varphi)$):

$$Y_{lm}(\theta,\varphi) = (-1)^m\sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}\, e^{im\varphi}L_{lm}(\cos\theta), \tag{A.35}$$

$$\int_0^{2\pi} d\varphi \int_0^{\pi} \sin\theta\, d\theta\, Y_{lm}(\theta,\varphi)\, Y^*_{l'm'}(\theta,\varphi) = \delta^{\mathcal{K}}_{ll'}\,\delta^{\mathcal{K}}_{mm'}, \tag{A.36}$$

$$L_l(\hat{\mathbf{n}}_1\cdot\hat{\mathbf{n}}_2) = \frac{4\pi}{2l+1}\sum_{m=-l}^{l} Y_{lm}(\hat{\mathbf{n}}_1)\, Y^*_{lm}(\hat{\mathbf{n}}_2), \tag{A.37}$$

$$Y^*_{lm} = Y_{l-m}, \quad Y_{lm}(-\hat{\mathbf{n}}) = (-1)^l Y_{lm}(\hat{\mathbf{n}}). \tag{A.38}$$

- Harmonic expansion of plane waves:

$$e^{i\,\mathbf{k}\cdot\mathbf{r}} = \sum_{l=0}^{\infty}(2l+1)\, i^l\, j_l(kr)\, L_l(\hat{\mathbf{n}}_{\mathbf{k}}\cdot\hat{\mathbf{n}}_{\mathbf{r}}) \tag{A.39}$$

The C_l's

Lemma: Let $f(\mathbf{x})$ be a homogeneous and isotropic random field. Then:

$$\langle a_{lm} a_{l'm'} \rangle \propto \delta^{\mathcal{K}}_{ll'}\, \delta^{\mathcal{K}}_{m-m'}. \tag{A.40}$$

Definition: For a homogeneous and isotropic field one defines the C_l's as

$$C_l = \langle |a_{lm}|^2 \rangle \tag{A.41}$$

Angular two-point correlation function

We define the angular two-point correlation function of a random field $f(\mathrm{x})$ as the two-point correlation function of its projection on the sky:

$$w(\hat{\mathbf{n}}_1, \hat{\mathbf{n}}_2) \equiv \langle f(\hat{\mathbf{n}}_1) f(\hat{\mathbf{n}}_2) \rangle. \tag{A.42}$$

If the field is isotropic this function must be invariant under $SO(3)$ transformations of $\hat{\mathbf{n}}_1$ and $\hat{\mathbf{n}}_2$, therefore it must only depend on $\cos\theta \equiv \hat{\mathbf{n}}_1 \cdot \hat{\mathbf{n}}_2$:

$$w(\theta) \equiv \langle f(\hat{\mathbf{n}}_1) f(\hat{\mathbf{n}}_2) \rangle \tag{A.43}$$

Lemma: For a homogeneous and isotropic field

$$\begin{aligned} w(\theta) &= \sum_{l=0}^{\infty} \frac{2l+1}{4\pi}\, C_l\, L_l(\cos\theta)) \\ C_l &= 2\pi \int_0^{\pi} \sin\theta\, d\theta\, L_l(\cos\theta))\, w(\theta) \end{aligned} \tag{A.44}$$

From $\mathbb{R}^3$ to $\mathcal{S}^2$

In order to relate three-dimensional quantities to quantities projected on the sphere, the harmonic expansion of plane waves (eq. (A.39)) can be used. Here are the most useful relations:

$$a_{lm} = \frac{4\pi\, i^l}{(2\pi)^{3/2}} \int_0^{\infty} k^2\, dk\, \lambda_l(k) \int d\Omega^2_{\mathbf{k}}\, Y_{lm}(\hat{\mathbf{n}}_{\mathbf{k}})\, f_{\mathbf{k}} \tag{A.45}$$

$$w(\theta) = \int_0^{\infty} \phi(z_1)\, dz_1 \int_0^{\infty} \phi(z_2)\, dz_2\, \xi(r(z_1, z_2, \theta)), \tag{A.46}$$

$$C_l = \frac{2}{\pi} \int k^2\, dk\, P_k \lambda_l^2(k) \tag{A.47}$$

where we have defined the functions

$$\lambda_l(k) \equiv \int_0^{\infty} \phi(z)\, j_l(k\, r(z))\, dz \tag{A.48}$$

and $r(z_1, z_2, \theta) \equiv [r^2(z_1) + r^2(z_2) - 2\, r(z_1)\, r(z_2)\, \cos\theta]^{-1/2}$. These relations can be extended to take into account redshift-space distortions [238, 239, 184].

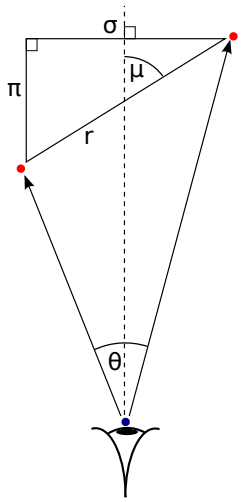

Figure A.1: Different coordinates used to describe the 3D anisotropic correlation functions

A.1.5 Redshift distortions and correlation functions

Redshift-space distortions (see section 2.1.3) may be due either to the alteration of the observed redshift by peculiar velocities or from using an incorrect cosmological model to translate redshifts into distance, and transform the real-space homogeneous and isotropic matter overdensity into a homogeneous but not isotropic field. Given the positions $\mathbf{x}_1$ and $\mathbf{x}_2$ of two points in space, it is useful to define the following set of coordinates (see figure A.1):

- θ: the angle subtended by both lines of sight:

$$\cos\theta \equiv \frac{\mathbf{x}_1 \cdot \mathbf{x}_2}{x_1\, x_2}. \tag{A.49}$$

- $r \equiv |\mathbf{x}_1 - \mathbf{x}_2|$: the 3D relative distance between both points

$$r \equiv \sqrt{x_1^2 + x_2^2 - 2\, x_1\, x_2\, \cos\theta}. \tag{A.50}$$

- μ: the angle of the separation vector with respect to the line of sight:

$$\mu \equiv \frac{(\mathbf{x}_1 - \mathbf{x_2}) \cdot (\mathbf{x}_1 + \mathbf{x_2})}{|\mathbf{x}_1 - \mathbf{x_2}|\, |\mathbf{x}_1 + \mathbf{x_2}|} = \frac{|x_1^2 - x_2^2|}{\sqrt{x_1^4 + x_2^4 + 2\, x_1^2\, x_2^2\, (1 - 2\, \cos^2\theta)}}. \tag{A.51}$$

- π: the projection of the separation vector along the line of sight:

$$\pi \equiv r\mu \equiv \frac{|x_1^2 - x_2^2|}{\sqrt{x_1^2 + x_2^2 + 2\, x_1\, x_2\, \cos\theta}}. \tag{A.52}$$

- σ: the projection of the separation vector on the plane transverse to the line of sight:

$$\sigma \equiv r\sqrt{1 - \mu^2} \equiv \frac{2\, x_1\, x_2\, \sin\theta}{\sqrt{x_1^2 + x_2^2 + 2\, x_1\, x_2\, \cos\theta}} \tag{A.53}$$

In direct analogy with the two pairs of coordinates (π, σ) and (r, μ) in configuration space, we define the Fourier space coordinates

$$k \equiv |\mathbf{k}|, \quad \mu_{\mathbf{k}} \equiv \mathbf{k} \cdot \hat{\mathbf{n}}_{\rm LOS}, \quad k_{\parallel} \equiv k\,\mu_{\mathbf{k}}, \quad k_{\perp} \equiv k\,\sqrt{1-\mu_{\mathbf{k}}^2}. \tag{A.54}$$

In redshift space and in the plane-parallel approximation, the observed power spectrum depends both on the modulus of the wave vector k and on its projection along the line of sight $\mu_{\mathbf{k}}$. $P(k, \mu_{\mathbf{k}})$ can be expanded in terms of Legendre polynomials of $\mu_{\mathbf{k}}$ (i.e. a multipole expansion):

$$P(k, \mu_{\mathbf{k}}) = \sum_{l \geq 0} P_l(k)\, L_l(\mu_{\mathbf{k}}). \tag{A.55}$$

The correlation function can also be expanded in multipoles $\xi(r, \mu) = \sum_l \xi_l(r)\, L_l(\mu)$, and the correlation function multipoles can be related to the $P_l(k)$s via

$$\xi_l(r) = \frac{i^l}{2\pi^2} \int_0^\infty P_l(k)\, j_l(kr)\, r^2\, dr. \tag{A.56}$$

In the linear approximation [78] the power spectrum multipoles are proportional to the real-space power spectrum

$$P_0(k) = \left(1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2\right) P_r(k), \tag{A.57}$$

$$P_2(k) = \left(\frac{4}{3}\beta + \frac{4}{7}\beta^2\right) P_r(k), \tag{A.58}$$

$$P_4(k) = \frac{8}{35}\beta^2\, P_r(k), \tag{A.59}$$

and higher multipoles are 0 (the RSD factor β was defined in eq. (2.24). The correlation function multipoles can also be related to the real-space 2PCF through:

$$\xi_0(r) = \left(1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2\right) \xi(r), \tag{A.60}$$

$$\xi_2(r) = \left(\frac{4}{3}\beta + \frac{4}{7}\beta^2\right) \left[\xi(r) - \bar{\xi}(r)\right], \tag{A.61}$$

$$\xi_4(r) = \frac{8}{35}\beta^2 \left[\xi(r) + \frac{5}{2}\bar{\xi}(r) - \frac{7}{2}\bar{\bar{\xi}}(r)\right], \tag{A.62}$$

$$\bar{\xi}(r) \equiv \frac{3}{r^3}\int_0^r \xi(s)\, s^2\, ds, \quad \bar{\bar{\xi}}(r) \equiv \frac{5}{r^5}\int_0^r \xi(s)\, s^4\, ds. \tag{A.63}$$

The anisotropic power spectrum and correlation function can also be directly related using the cylindrical Bessel function J_0:

$$\xi(\pi, \sigma) = \frac{1}{4\pi^2}\int_{-\infty}^{\infty} dk_{\parallel}\, e^{i\,k_{\parallel}\pi} \int_0^\infty dk_{\perp}\, k_{\perp}\, J_0(k_{\perp}\sigma)\, P(k_{\parallel}, k_{\perp}). \tag{A.64}$$

A.2 Discrete tracers of the matter distribution

In order to study the properties and evolution of random fields in cosmology, often discrete tracers are used, such as galaxies. Here we will give a mathematical description of these point processes and discuss the techniques that can be used to extract information about the underlying random field. Much of the information contained within this section has been extracted from the references [240, 150]. Only the results that have been used for this thesis are discussed here, and the reader is encouraged to consult these references for further details.

A.2.1 Point processes

Definition

A point process Φ in $\mathbb{R}^3$ is a random variable taking values in the family of discrete subsets of $\mathbb{R}^3$, $\Phi \to \phi \equiv \{\mathbf{x}_i \in \mathbb{R}^3\}$. The realizations ϕ of Φ must satisfy two conditions:

- **Local finiteness:** any bound subset of $\mathbb{R}^3$ contains a finite number of points.
- **Simplicity:** $\mathbf{x}_i \neq \mathbf{x}_j \ \forall\, \mathbf{x}_i, \mathbf{x}_j \in \phi$.

For a point process Φ and a subset $A \subset \mathbb{R}^3$, we can define another random variable $\Phi(A)$, defined as the number of points of the process contained in A. The fact that we use the same label Φ for both random variables is intentional, since both fully describe the statistical nature of the same object.

Intensity functions

Let Φ be a point process. The first-order intensity function of Φ, $\lambda_1(\mathbf{x})$ is defined as the local density of points averaged over all realizations of the process:

$$\lambda_1(\mathbf{x}) \equiv \lim_{dV \to 0} \frac{\langle \Phi(dV) \rangle}{dV}, \tag{A.65}$$

where dV is a region of $\mathbb{R}^3$ containing $\mathbf{x}$. Likewise we can define the N-point intensity function $\lambda_N(\mathbf{x}_1, ..., \mathbf{x}_N)$:

$$\lambda_N(\mathbf{x}_1, ..., \mathbf{x}_N) \equiv \lim_{dV_i \to 0} \frac{\langle \prod_i \Phi(dV_i) \rangle}{\prod_i dV_i}. \tag{A.66}$$

Poisson processes

A Poisson point process satisfies the following two conditions:

- $\Phi(\varnothing) = 0$.
- The number of points in disjoint subsets are independent of each other:

$$P(\Phi(A), \Phi(B)) = P(\Phi(A))\, P(\Phi(B)), \quad \text{for } A \cap B = \varnothing \tag{A.67}$$

- The number of points in a given subset A depends only on the integral of the first-order intensity function over A:

$$\Lambda(A) \equiv \int_A \lambda_1(\mathbf{x})\, dV. \tag{A.68}$$

It can be proven that the probability distribution for a Poisson process is

$$P(\Phi(A) = n) = \frac{[\Lambda(A)]^n}{n!} e^{-\Lambda(A)}. \tag{A.69}$$

The galaxy distribution can be understood as one realization of a three-dimensional point process. This point process can be connected with the underlying continuous galaxy number density field $n(\mathbf{x}) = \bar{n}(1 + \delta(\mathbf{x}))$ modelling it through a Poisson process with intensity $\lambda_1 \equiv n$ (although this is not the only possible description). Since, as we described at the beginning of this appendix, the overdensity δ is itself a random field, this is a doubly stochastic process, also referred to as a Cox process.

Correlation functions

Consider a Poisson point process with intensity λ. The probability of finding two objects within a volumes dV_1 and dV_2 of two points $\mathbf{x}_1$ and $\mathbf{x}_2$ is

$$dP(\mathbf{x}_1, \mathbf{x}_2) = \left[\lambda(\mathbf{x}_1)\lambda(\mathbf{x}_2) + \lambda(\mathbf{x}_1)\,\delta^{\mathcal{D}}(\mathbf{x}_1 - \mathbf{x}_2)\right]\, dV_1\, dV_2 + \mathcal{O}(dV^3). \tag{A.70}$$

In the following section we will describe different techniques to calculate the galaxy correlation function, all of which avoid considering galaxy self-pairs, so we can safely ignore the second term $\propto \delta^{\mathcal{D}}(\mathbf{x}_1 - \mathbf{x}_2)$ above. Taking the ensemble average over the realizations of the intensity field, the probability of finding two galaxies at $\mathbf{x}_1$ and $\mathbf{x}_2$ is directly related to the 2PCF of the random field:

$$\langle dP(\mathbf{x}_1, \mathbf{x}_2)\rangle = \bar{n}^2\,[1 + \xi(\mathbf{x}_1, \mathbf{x}_2)]\, dV_1\, dV_2. \tag{A.71}$$

A.2.2 Estimating the correlation function from discrete data

As we have just seen, the 2PCF $\xi(r)$ of a random field can be understood as the average excess probability, with respect to a random homogeneous point distribution, of finding two points of the associated Cox process separated by a distance r (in this section we will assume that the field is homogeneous and isotropic, so the correlation function only depends on r. The results presented here are easily generalizable to other cases). Thus, we can estimate the correlation function from the point process as

$$1 + \hat{\xi}(r) = \frac{\#\text{ pairs in the distribution at a distance } r}{\#\text{ pairs expected for a random distribution}} \tag{A.72}$$

This is, in principle, not so difficult to compute, and can be written as

$$1 + \hat{\xi}(r) = \frac{1}{N_D}\sum_{i}^{N_D} \frac{\sum_{j\neq i}\Theta(r < |\mathbf{x}_i - \mathbf{x_j}| < r + dr)}{\bar{n}\,V(r, dr)}, \tag{A.73}$$

where $V(r, dr)$ is the volume of a spherical shell of radius r and width dr.

This estimator is perfectly fine as long as we can draw spheres of radius r around all the galaxies in our dataset. This is the case, for example, for data coming from an N-body simulation, in which the boundary conditions are periodic (see figure A.2), however, in realistic cases the data lies inside a limited region of space, and spheres drawn around objects which are close to the boundaries will lie partly outside the region. For these objects, dividing by the full volume of these spheres in eq. (A.73) will overestimate the expected number of random objects, biasing the estimation of ξ. This situation can be remedied by using the volume enclosed within the observed region $V(\mathbf{x}_i, r, dr)$ instead of the full volume. However this quantity may be extremely difficult (even impossible) to calculate analytically for general boundary shapes. Even worse is the situation in which some regions have been mapped differently, so that the number density of observed objects varies artificially in space, as is usually the case in galaxy surveys.

These caveats are all related to the same problem: how can we calculate with good precision the expected number of pairs if the point distribution were homogeneous? Fortunately this question is also easy to answer: generate a random homogeneous distribution with the same characteristics as your data and calculate this quantity directly from that distribution. This solution has given rise to a whole family of methods to calculate the 2PCF using random distribution, the simplest of which is:

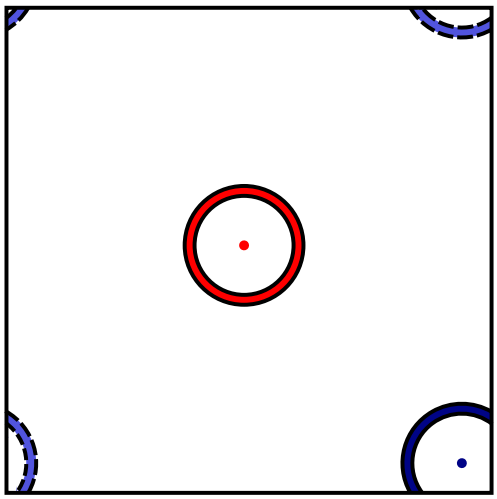

Figure A.2: Boundary issues related to the calculation of two-point correlation functions. Spheres can be drawn around points far away from the boundaries (red). For points close to the boundaries, the volume of the intersection of the sphere and the surveyed region must be calculated (dark blue), for which a Montecarlo integration via a random catalogue is usually the best solution. This is not an issue for periodic boundary conditions (light blue).

1. Calculate the number of pairs in your data separated by a distance r. We shall call this histogram $dd(r)$.

2. Generate a random dataset with N_R objects that includes any artificial effect that may be present the data (selection function, mask, completeness, etc.).

3. Calculate the same histogram for the random catalogue $rr(r)$.

4. Scale both histograms by the total number of pairs in each catalogue

$$DD(r) \equiv \frac{2}{N_D\,(N_D-1)}\,dd(r), \quad RR(r) \equiv \frac{2}{N_R\,(N_R-1)}\,rr(r). \tag{A.74}$$

   This step is necessary, since a larger number of random objects is needed in order to suppress the Poisson variance associated to the random catalogue.

5. Estimate the correlation function as

$$\hat{\xi}_{\rm PH}(r) \equiv \frac{DD(r)}{RR(r)} - 1. \tag{A.75}$$

Similar estimators have been proposed in the literature, often making use of data-random cross correlated pair counts $DR(r)$, which can be used to deal more efficiently with boundary corrections. In 1993 Landy and Szalay [176] studied the bias and variance of several of these, and worked out the optimal estimator for a random distribution. This is the Landy & Szalay estimator, given by

$$\hat{\xi}_{\rm LS}(r) \equiv \frac{DD(r) - 2\,DR(r) + RR(r)}{RR(r)}. \tag{A.76}$$

Recently, this study has been extended to find the optimal estimator also for weakly clustered distributions [242]. The main advantage of this approach over the Landy & Szalay estimator

Estimator	Properties
$\hat{\xi}_{\rm PH} = \frac{DD}{RR} - 1$	Unbiased, large variance
$\hat{\xi}_{\rm DP} = \frac{DD}{DR} - 1$	Small bias, large variance
$\hat{\xi}_{\rm Ham} = \frac{DD\,RR}{DR^2} - 1$	Small bias, small variance
$\hat{\xi}_{\rm LS} = \frac{DD - 2\,DR + RR}{RR}$	Unbiased, small variance

Table A.1: Summary of different estimators for the 2PCF. The performance of these was explored in [176, 241]. Both the Landy & Szalay and Hamilton estimators outperform the rest in terms of variance and ability to cope with boundary effects.

is that fewer random objects must be generated to obtain the same accuracy, however, this estimator needs to be calibrated first by calculating the pair histograms in several mock catalogues. Table A.1 shows a summary of the different estimators proposed in the literature.

As we have said, the key to a good estimation of the 2PCF is usually to introduce all artificial effects that may bias the estimator in the random catalogue. However sometimes it may be easier or more convenient to add different weights to galaxies in different regions in space to compensate for these effects (this is done, for instance, to solve the problem of obscuration by stars in galaxy surveys [188]). The possibility to add weights to galaxies can also be used to improve the variance of the estimator (e.g. [243]).

Finally, it is worth pointing out that, since calculating the 2PCF involves counting pairs of objects, the computational time for this task grows very fast, with the square of the number of objects. For this reason, for large datasets it is necessary to use alternative algorithms and parallel computing tools. The references [244, 245, 246, 247] illustrate some approaches to this problem.

APPENDIX B

Gaussian covariance matrices for galaxy correlation functions

The cosmological constraints drawn from any analysis of the correlation function depend strongly on the estimated statistical uncertainties of this observable. Since, in cosmology, we only have access to one "realization" of the Universe, it is not possible to perform the same experiment several times so that uncertainties can be estimated from ensemble averages. This, and the different methods that can be used to estimate the covariance matrix of the 2PCF have already been discussed in section 3.1.3. It is possible to obtain a fast theoretical estimate of the covariance matrix if we assume that the observed overdensity field is Gaussianly distributed. It is difficult to include every observational effect that may influence the statistical errors in this calculation, however they are a useful tool as an order-of-magnitude estimate, and are frequently used to forecast the cosmological constraints that a given experiment may be able to obtain. See the references [163, 248, 249] for more ideas on this type of estimations. Throughout this section we will use the following convention for the power spectrum:

$$\langle \delta_{\mathbf{k}} \delta_{\mathbf{k}'} \rangle = \delta^{\mathcal{D}}(\mathbf{k} + \mathbf{k}') P(\mathbf{k}) = \frac{V}{(2\pi)^3} \delta^{\mathcal{K}}_{\mathbf{k},-\mathbf{k}'} P(\mathbf{k}), \tag{B.1}$$

where V is the volume of your "box" (which we identify with the surveyed volume). The fact that the power spectrum and the correlation function are measured from a discrete dataset introduces a Poisson uncertainty (shot noise), which boils down to the contribution:

$$P(\mathbf{k}) \to P(\mathbf{k}) + \frac{1}{\bar{n}}, \tag{B.2}$$

where $\bar{n}$ is the 3D comoving number density of sources in the survey. This extra term has its origin in the Poisson contribution in equation A.70.

Figure B.1 shows some examples the correlation matrices estimated with the approach detailed here.

B.1 Correlation function in real space

Since, in this case, the power spectrum depends only on $k \equiv |\mathbf{k}|$, given a realization of the Fourier-space overdensity $\delta_{\mathbf{k}}$ in a region of volume V, $P_r(k)$ can be estimated by averaging

over all modes with the same value of k:

$$\hat{P}_r(k) \equiv \frac{1}{4\pi}\int d\Omega_k \left(\frac{(2\pi)^3}{V}|\delta_{\mathbf{k}}|^2 - \frac{1}{\bar{n}}\right), \tag{B.3}$$

where $d\Omega_k$ is the solid angle element in k-space. Notice that we have subtracted the Poisson term. The covariance matrix for this estimator is

$$\begin{aligned} C^P(k_1,k_2) &\equiv \langle \hat{P}_r(k_1)\hat{P}_r(k_2)\rangle - \langle \hat{P}_r(k_1)\rangle\langle \hat{P}_r(k_2)\rangle \\ &= \frac{(2\pi)^6}{(4\pi)^2 V^2}\int d\Omega_{k_1}\int d\Omega_{k_2}\left(\langle |\delta_{\mathbf{k_1}}|^2|\delta_{\mathbf{k_2}}|^2\rangle - \langle |\delta_{\mathbf{k_1}}|^2\rangle\langle |\delta_{\mathbf{k_2}}|^2\rangle\right). \end{aligned}$$

We are assuming that the overdensity field has Gaussian statistics and therefore that Wick's theorem holds:

$$\langle \delta_1\delta_2\delta_3\delta_4\rangle = \langle \delta_1\delta_2\rangle\langle \delta_3\delta_4\rangle + \langle \delta_1\delta_3\rangle\langle \delta_2\delta_4\rangle + \langle \delta_1\delta_4\rangle\langle \delta_2\delta_3\rangle. \tag{B.4}$$

Using this together with eqs. (B.1) and (B.2) it is easy to obtain the result:

$$C^P(k_1,k_2) = \frac{4\,\pi^2}{V}\frac{\delta^{\mathcal{D}}(k_1-k_2)}{k_1^2}\left(P_r(k_1)+\frac{1}{\bar{n}}\right)^2. \tag{B.5}$$

Note that usually the number density is not constant, but a function of redshift. A way to address this is to interpret the combination $(P_r(k)+1/\bar{n})^2/V$ as the inverse variance integrated over the survey volume if the number density were constant. Hence we can extend this to a z-dependent $\bar{n}$ defining:

$$\mathfrak{P}^{-2}(k) \equiv \int \frac{dV}{[P_r(k)+1/\bar{n}]^2}, \tag{B.6}$$

where the integral covers the whole survey volume. Thus the covariance matrix is

$$C^P(k_1,k_2) = 4\,\pi^2\frac{\delta^{\mathcal{D}}(k_1-k_2)}{k_1^2}\mathfrak{P}^2(k_1). \tag{B.7}$$

Now, relating the power spectrum to the correlation function it is easy to find a similar expression for the covariance matrix of the latter [249]:

$$C^\xi(r_1,r_2) = \frac{1}{\pi^2}\int_0^\infty dk\,k^2\,j_0(kr_1)\,j_0(kr_2)\mathfrak{P}^2(k). \tag{B.8}$$

Note that this would yield the covariance of the correlation function calculated from a finite Poisson realization in infinitesimal intervals (bins) of r. Unless we take into account the finiteness of the bins, the shot noise term will give an infinite contribution in the diagonal of this matrix, since:

$$\int_0^\infty dk\,k^2\,j_0(kr_1)\,j_0(kr_2) = \frac{\pi}{2}\frac{\delta^{\mathcal{D}}(r_1-r_2)}{r_1^2}.$$

It is possible to take into account the finite size Δr of the bins considering that what we actually observe is the average of the 2PCF in each bin

$$\xi_r(r;\Delta r) \equiv \frac{3}{r_2^3-r_1^3}\int_{r_1}^{r_2} dr\,r^2\,\xi_r(r), \tag{B.9}$$

where $r_1 = r - \Delta r/2$ and $r_2 = r + \Delta r/2$. This reflects on the previous formula by regularizing the Bessel function:

$$j_l(kr) \to \bar{j}_l(kr; \Delta r) \equiv \frac{3}{x_2^3 - x_1^3} \int_{x_1}^{x_2} dx\, x^2\, j_l(x), \tag{B.10}$$

with $x_i \equiv k\, r_i$. In the case $l = 0$ this simplifies to

$$\bar{j}_0(kr; \Delta r) \equiv 3\, \frac{x_2^2\, j_1(x_2) - x_1^2\, j_1(x_1)}{x_2^3 - x_1^3}. \tag{B.11}$$

We will omit the Δr from now on. Taking all this into account the final result is

$$C^\xi(r_1, r_2) = \frac{1}{\pi^2} \int_0^\infty dk\, k^2\, \bar{j}_0(kr_1)\, \bar{j}_0(kr_2) \mathfrak{P}^2(k). \tag{B.12}$$

B.2 Redshift space in the (π, σ) frame

In the presence of redshift-space distortions, the only symmetry remaining in k-space corresponds to the azimuthal angle. Thus we can still estimate the anisotropic power spectrum as an average over this coordinate:

$$\hat{P}(k_\parallel, k_\perp) \equiv \frac{1}{2\pi} \int_0^{2\pi} d\phi_k \left(\frac{(2\pi)^3}{V} |\delta_{\mathbf{k}}|^2 - \frac{1}{\bar{n}} \right) \tag{B.13}$$

Following the same steps as before, the covariance matrix for this estimator can be calculated as

$$C^P(k_{1,\parallel}, k_{1,\perp}; k_{2,\parallel}, k_{2,\perp}) = 4\pi^2 \frac{\delta^{\mathcal{D}}(k_{2,\perp} - k_{1,\perp})}{k_{1,\perp}} \left[\delta^{\mathcal{D}}(k_{2,\parallel} - k_{1,\parallel}) + \delta^{\mathcal{D}}(k_{2,\parallel} + k_{1,\parallel}) \right] \mathfrak{P}^2(k_\parallel, k_\perp),$$

where $\mathfrak{P}^2(k_\parallel, k_\perp)$ has the same definition as before.

Assuming bin sizes $\Delta\pi$ and $\Delta\sigma$, the covariance matrix for the anisotropic correlation function can be calculated as

$$C^\xi(\pi_1, \sigma_1; \pi_2, \sigma_2) \equiv \frac{1}{\pi^2} \int_0^\infty dk_\parallel\, \bar{\cos}(k_\parallel \pi_1)\, \bar{\cos}(k_\parallel \pi_2) \int_0^\infty dk_\perp\, k_\perp\, \bar{J}_0(k_\perp \sigma_1)\, \bar{J}_0(k_\perp \sigma_2)\, \mathfrak{P}^2(k_\parallel, k_\perp).$$

Here $\bar{\cos}$ and $\bar{J}_0$ are the regularized versions of cos and J_0 defined as

$$\bar{\cos}(k_\parallel \pi) \equiv \frac{\sin(x_2) - \sin(x_1)}{x_2 - x_1}, \quad (x_{1,2} \equiv k_\parallel(\pi \pm \Delta\pi)) \tag{B.14}$$

$$\bar{J}_0(k_\perp \sigma) \equiv 2\, \frac{x_2\, J_1(x_2) - x_1\, J_1(x_1)}{x_2^2 - x_1^2}, \quad (x_{1,2} \equiv k_\perp(\sigma \pm \Delta\sigma)) \tag{B.15}$$

B.2.1 Particular case: the radial correlation function

In the case $\sigma_{1,2} = 0$, the previous result simplifies further:

$$C_\parallel^\xi(\pi_1, \pi_2) \equiv \frac{1}{\pi^2} \int_0^\infty dk_\parallel\, \bar{\cos}(k_\parallel \pi_1)\, \bar{\cos}(k_\parallel \pi_2)\, \left[\mathfrak{P}_\parallel(k_\parallel, \Delta\sigma) \right]^2, \tag{B.16}$$

where we have defined the projected k-space variance

$$\left[\mathfrak{P}_\parallel(k_\parallel, \Delta\sigma) \right]^2 \equiv \int_0^\infty dk_\perp\, k_\perp \left[2 \frac{J_1(k_\perp\, \Delta\sigma)}{k_\perp\, \Delta\sigma} \right]^2 \mathfrak{P}^2(k_\parallel, k_\perp) \tag{B.17}$$

B.3 Redshift space in the (r, μ) frame

We can estimate the l-th multipole of the power spectrum from a realization of the $\delta_{\mathbf{k}}$ as

$$\hat{P}_l(k) \equiv \frac{2l+1}{4\pi} \int d\Omega_{\mathbf{k}}^2 \left(\frac{(2\pi)^3}{V} |\delta_{\mathbf{k}}|^2 - \frac{1}{\bar{n}} \right) L_l(\mu_k). \tag{B.18}$$

Following the same steps as before, the covariance matrix for this estimator is

$$C^P(l_1, k_1; l_2, k_2) = \frac{4\pi^2}{V} \frac{\delta^{\mathcal{D}}(k_1 - k_2)}{k_1^2} \frac{(2l_1+1)(2l_2+1)}{2} \int_{-1}^{1} d\mu_k \, L_{l_1}(\mu_k) \, L_{l_2}(\mu_k) \left(P(k, \mu_k) + \frac{1}{\bar{n}} \right)^2.$$

Expanding $P(k, \mu_k)$ in multipoles this is:

$$C^P(l_1, k_1; l_2, k_2) = \frac{4\pi^2}{V} \frac{\delta^{\mathcal{D}}(k_1 - k_2)}{k_1^2} \left[\sum_{l_3, l_4} B_{l_1 \, l_2 \, l_3 \, l_4} P_{l_3}(k) P_{l_4}(k) + 2 \sum_{l_3} C_{l_1 \, l_2 \, l_3} \frac{P_{l_3}(k)}{\bar{n}} + (2l_1+1) \frac{\delta^{\mathcal{K}}_{l_1, l_2}}{\bar{n}^2} \right],$$

where

$$B_{l_1 \, l_2 \, l_3 \, l_4} \equiv \frac{(2l_1+1)(2l_2+1)}{2} \int_{-1}^{1} d\mu \, L_{l_1}(\mu) \, L_{l_2}(\mu) \, L_{l_3}(\mu) \, L_{l_4}(\mu)$$

$$C_{l_1 \, l_2 \, l_3} \equiv \frac{(2l_1+1)(2l_2+1)}{2} \int_{-1}^{1} d\mu \, L_{l_1}(\mu) \, L_{l_2}(\mu) \, L_{l_3}(\mu) = (2l_1+1)(2l_2+1) \begin{pmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{pmatrix}.$$

This can be simplified further in the Kaiser approximation. In this case the only non-zero multipoles are $l = 0$, 2, 4 and they can be written as $P_l(k) = v_l(\beta) \, P_r(k)$. Here

$$v_0(\beta) = 1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2, \;\; v_2(\beta) = \frac{4}{3}\beta + \frac{4}{7}\beta^2, \;\; v_4(\beta) = \frac{8}{35}\beta^2. \tag{B.19}$$

Then the previous expression can be written as

$$C^P(l_1, k_1; l_2, k_2) = \frac{4\pi^2}{V} \frac{\delta^{\mathcal{D}}(k_1 - k_2)}{k_1^2} \left[\mathcal{A}_{l_1, l_2}(\beta) P_r^2(k_1) + 2\mathcal{B}_{l_1, l_2}(\beta) \frac{P_r(k_1)}{\bar{n}} + (2l_1+1) \frac{\delta^{\mathcal{K}}_{l_1, l_2}}{\bar{n}^2} \right], \tag{B.20}$$

where

$$\mathcal{A}_{l_1, l_2}(\beta) \equiv \sum_{l_3, l_4} B_{l_1 \, l_2 \, l_3 \, l_4} v_{l_3}(\beta) v_{l_4}(\beta), \quad \mathcal{B}_{l_1, l_2}(\beta) \equiv \sum_{l_3} C_{l_1 \, l_2 \, l_3} v_{l_3}(\beta). \tag{B.21}$$

Now, defining the variance

$$\mathfrak{P}^{-2}_{l_1, l_2}(k) \equiv \int dV \left[\mathcal{A}_{l_1, l_2}(\beta) P_r^2(k) + 2\mathcal{B}_{l_1, l_2}(\beta) \frac{P_r(k)}{\bar{n}(z)} + (2l_1+1) \frac{\delta^{\mathcal{K}}_{l_1, l_2}}{\bar{n}^2(z)} \right]^{-1} \tag{B.22}$$

to account for a non-homogeneous number density, the covariance matrix can be written as

$$C^P(l_1, k_1; l_2, k_2) = 4\pi^2 \frac{\delta^{\mathcal{D}}(k_1 - k_2)}{k_1^2} \mathfrak{P}^2_{l_1, l_2}(k). \tag{B.23}$$

As before, we can now compute the covariance matrix for the correlation function multipoles:

$$C^{\xi}(l_1, r_1; l_2, r_2) = \frac{i^{l_1 + l_2}}{\pi^2} \int_0^{\infty} dk \, k^2 \, \bar{j}_{l_1}(kr_1) \, \bar{j}_{l_2}(kr_2) \, \mathfrak{P}^2_{l_1, l_2}(k). \tag{B.24}$$

For completeness, the values of $\mathcal{A}_{l_1,l_2}(\beta)$ for $l_{1,2} \leq 4$ in the Kaiser approximation are

$$\begin{aligned}
\mathcal{A}_{0,0}(\beta) &= 1 + \frac{4}{3}\beta + \frac{6}{5}\beta^2 + \frac{4}{7}\beta^3 + \frac{1}{9}\beta^4 \\
\mathcal{A}_{0,2}(\beta) = \mathcal{A}_{2,0}(\beta) &= \frac{8}{3}\beta + \frac{24}{7}\beta^2 + \frac{40}{21}\beta^3 + \frac{40}{99}\beta^4 \\
\mathcal{A}_{0,4}(\beta) = \mathcal{A}_{4,0}(\beta) &= \frac{48}{35}\beta^2 + \frac{96}{77}\beta^3 + \frac{48}{143}\beta^4 \\
\mathcal{A}_{2,2}(\beta) &= 5 + \frac{220}{21}\beta + \frac{90}{7}\beta^2 + \frac{1700}{231}\beta^3 + \frac{2075}{1287}\beta^4 \\
\mathcal{A}_{2,4}(\beta) = \mathcal{A}_{4,2}(\beta) &= \frac{48}{7}\beta + \frac{816}{77}\beta^2 + \frac{6960}{1001}\beta^3 + \frac{240}{143}\beta^4 \\
\mathcal{A}_{4,4}(\beta) &= 9 + \frac{1404}{77}\beta + \frac{104166}{5005}\beta^2 + \frac{11772}{1001}\beta^3 + \frac{6399}{2431}\beta^4.
\end{aligned}$$

and the $\mathcal{B}_{l_1,l_2}$'s are

$$\begin{aligned}
\mathcal{B}_{0,0}(\beta) &= 1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2 \\
\mathcal{B}_{0,2}(\beta) = \mathcal{B}_{2,0}(\beta) &= \frac{4}{3}\beta + \frac{4}{7}\beta^2 \\
\mathcal{B}_{0,4}(\beta) = \mathcal{B}_{4,0}(\beta) &= \frac{8}{35}\beta^2 \\
\mathcal{B}_{2,2}(\beta) &= 5 + \frac{110}{21}\beta + \frac{15}{7}\beta^2 \\
\mathcal{B}_{2,4}(\beta) = \mathcal{B}_{4,2}(\beta) &= \frac{24}{7}\beta + \frac{136}{77}\beta^2 \\
\mathcal{B}_{4,4}(\beta) &= 9 + \frac{702}{77}\beta + \frac{17361}{5005}\beta^2
\end{aligned}$$

B.3.1 Special case: the monopole

Now, for $l_1 = l_2 = 0$ we have:

$$B_{0\,0\,l_3\,l_4} = \frac{\delta^{\mathcal{K}}_{l_3,l_4}}{2l_3+1}, \quad C_{0\,0\,l_3} = \delta^{\mathcal{K}}_{l_3,0}, \tag{B.25}$$

$$\mathcal{A}_{0\,0}(\beta) = \sum_l \frac{v_l^2(\beta)}{2l+1}, \quad \mathcal{B}_{0\,0}(\beta) = v_0(\beta), \tag{B.26}$$

$$\mathfrak{P}^{-2}_{0,0}(k) = \int dV \left[\left(v_0(\beta)P_r(k) + \frac{1}{\bar{n}(z)}\right)^2 + P_r^2(k)\sum_{l=1}\frac{v_l^2(\beta)}{2l+1}\right]^{-1}. \tag{B.27}$$

Note that the last term in the integrand for $\mathfrak{P}^{-2}$ above is missing from previous works.

The covariance is

$$C^{\xi}_0(r_1,r_2) = \frac{1}{\pi^2}\int_0^\infty dk\, k^2\, \bar{j}_0(kr_1)\,\bar{j}_0(kr_2)\,\mathfrak{P}^2_{0,0}(k). \tag{B.28}$$

B.4 Angular covariance

For the overdensity projected on the sphere $\delta(\hat{\mathbf{n}})$ defined in equation (A.26) the coefficients a_{lm} are the analogue in S^2 of the $\delta_{\mathbf{k}}$'s in $\mathbb{R}^3$. As in the three-dimensional case, when the

angular correlation function is computed from a discrete dataset, the variance of the a_{lm}'s receives a shot noise contribution:

$$C_l \to C_l + \frac{1}{\bar{\sigma}_n}, \tag{B.29}$$

where $\bar{\sigma}_n$ is the two-dimensional number density.

As in the 3D case, the C_l's can be estimated from the a_{lm}'s by averaging over the symmetric m modes:

$$\hat{C}_l \equiv \frac{1}{2l+1} \sum_{m=-l}^{l} \left(|a_{lm}|^2 - \frac{1}{\bar{\sigma}_n} \right), \tag{B.30}$$

The variance of this estimator, assuming the a_{lm}'s are Gaussian, is:

$$C^C(l_1, l_2) = \frac{2}{2l_1+1} \left(C_{l_1} + \frac{1}{\bar{\sigma}_n} \right)^2 \delta^{\mathcal{K}}_{l_1,l_2}. \tag{B.31}$$

With this it is easy to derive the variance of the angular correlation function:

$$C^w(\theta_1, \theta_2) = \frac{2}{4\pi} \sum_{l \leq 0} \frac{(2l+1)}{4\pi} \left(C_l + \frac{1}{\bar{\sigma}_n} \right)^2 L_l(\cos\theta_1)\, L_l(\cos\theta_2) \tag{B.32}$$

As before, the finiteness of the bins in θ can be addressed by using the regularized Legendre polynomials, defined as:

$$\bar{L}_l(\cos\theta) \equiv \frac{1}{x_1 - x_2} \int_{x_2}^{x_1} L_l(x)\, dx = \frac{[L_{l+1}(x_1) - L_{l-1}(x_1)] - [L_{l+1}(x_2) - L_{l-1}(x_2)]}{(2l+1)(x_1 - x_2)}, \tag{B.33}$$

where $x_{1,2} = \cos(\theta \pm \Delta\theta/2)$.

Also, these results are exact assuming that our data covers the whole 4π solid angle. For galaxy surveys this is never the case, and only pseudo-multipoles can be estimated, which are not uncorrelated. However, it has been shown [163] that an incomplete sky coverage can be accounted for empirically by scaling this covariance by the available fraction of the sky $f_{\rm sky}$. Thus, the final expression for the covariance of the angular correlation function is:

$$C^w(\theta_1, \theta_2) \equiv \frac{2}{4\pi\, f_{\rm sky}} \sum_{l \leq 0} \frac{(2l+1)}{4\pi} \left(C_l + \frac{1}{\bar{\sigma}_n} \right)^2 \bar{L}_l(\cos\theta_1)\, \bar{L}_l(\cos\theta_2) \tag{B.34}$$

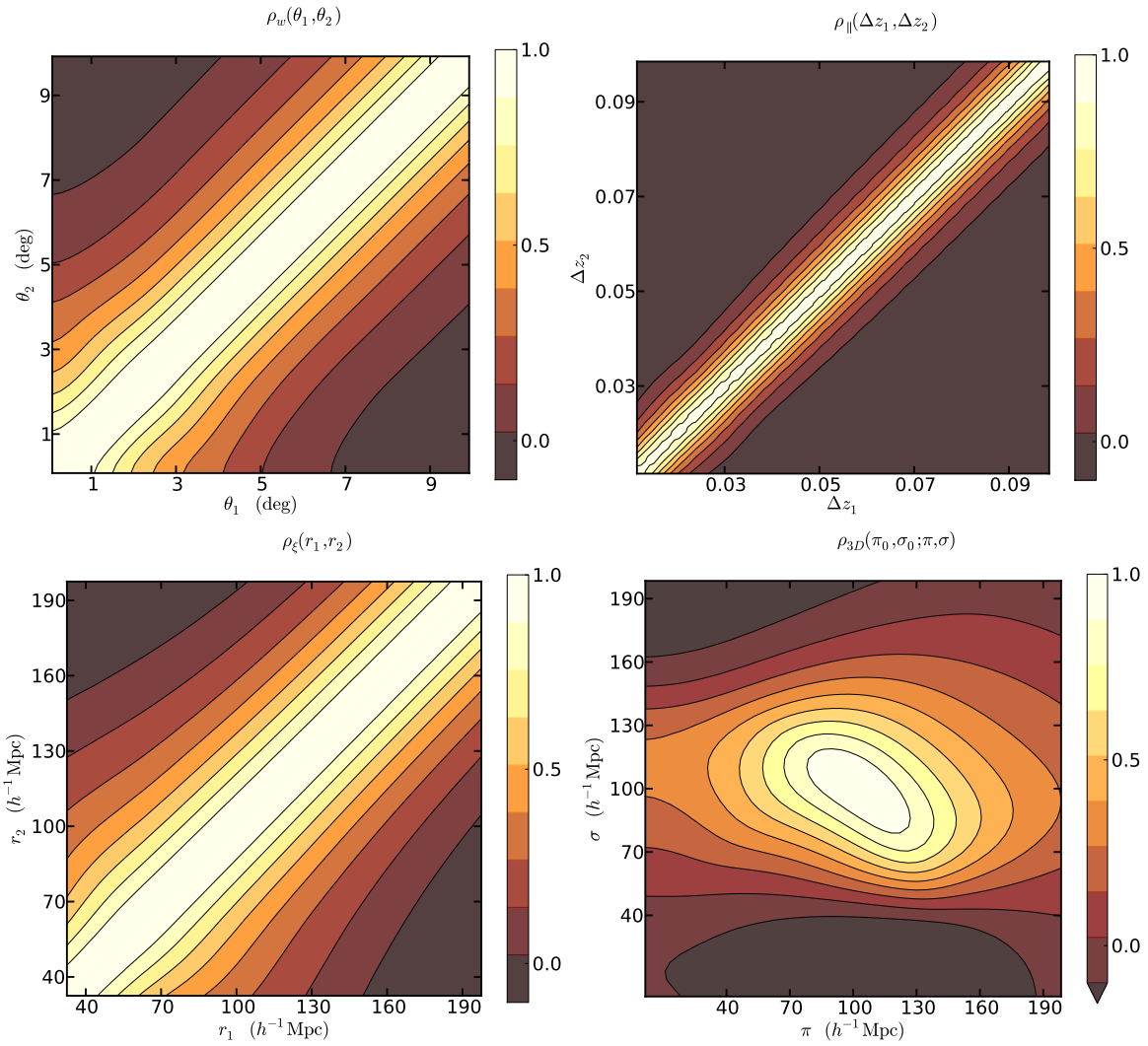

Figure B.1: Gaussian correlation matrices $\rho_{ij} = C_{ij}/\sqrt{C_{ii}\,C_{jj}}$ for the angular correlation function (top left panel), the radial correlation function (top right), the monopole (bottom left) and the anisotropic correlation function at the fixed point $\pi_0 = \sigma_0 = 100\,\mathrm{Mpc}/h$ (bottom right).

APPENDIX C

Geodesics and curvature tensors of the LTB metric

C.1 Geodesics

The geodesic equations ($\ddot{x}^\mu + \Gamma^\mu_{\nu\sigma}\dot{x}^\nu\dot{x}^\sigma = 0$) are:

$$\ddot{t} = -\frac{A'\dot{A}'}{1-k(r)}\dot{r}^2 - A\,\dot{A}(\dot{\theta}^2 + \sin^2\theta\dot{\varphi}^2), \tag{C.1}$$

$$\ddot{r} = -2\frac{\dot{A}'}{A'}\dot{r}\dot{t} - \frac{k'(r)A' + 2[1-k(r)]A''}{2[1-k(r)]A'}\dot{r}^2 + \frac{A[1-k(r)]}{A'}(\dot{\theta}^2 + \sin^2\theta\dot{\varphi}^2), \tag{C.2}$$

$$\ddot{\theta} = -2\frac{\dot{A}}{A}\dot{t}\dot{\theta} - 2\frac{A'}{A}\dot{r}\dot{\theta} + \cos\theta\,\sin\theta\,\dot{\varphi}^2, \tag{C.3}$$

$$\frac{d}{d\lambda}(A^2\sin^2\theta\dot{\varphi}) = 0. \tag{C.4}$$

Here a $\dot{()}$ over a coordinate (t, r, θ, φ) implies differentiation with respect to an affine parameter λ of the curve, while it denotes ∂_t when it is over $A(t, r)$.

Radial geodesics. Let us first study the case when the observer is at the centre of symmetry $r = 0$ and the geodesic passes through that point. Substituting $A(r = 0, t) = 0$ into (C.4) we obtain

$$A^2\sin^2\theta\dot{\varphi} = \text{const.} = A^2\sin^2\theta\dot{\varphi}\big|_{r=0} = 0, \tag{C.5}$$

therefore $\varphi = \text{const.}$. This can be substituted into (C.3) to yield

$$\frac{1}{A^2\dot{\theta}}\frac{d}{d\lambda}(A^2\dot{\theta}) = 0. \tag{C.6}$$

With the same trick as before, this implies $\theta = \text{const.}$. Hence, we are left with

$$\ddot{t} = -\frac{A'\dot{A}'}{1-k(r)}\dot{r}^2, \quad \ddot{r} = -2\frac{\dot{A}'}{A'}\dot{r}\dot{t} - \frac{k'(r)A' + 2[1-k(r)]A''}{2[1-k(r)]A'}\dot{r}^2. \tag{C.7}$$

The latter can be rewritten in terms of the coordinate

$$\chi \equiv \int^r \frac{dr}{\sqrt{1-k(r)}}, \tag{C.8}$$

to obtain

$$\ddot{t} + A'\dot{A}'\dot{\chi}^2 = 0, \quad \frac{1}{A'^2}\frac{d}{d\lambda}(A'^2\dot{\chi}) = \frac{\partial^2_\chi A}{\partial_\chi A}\dot{\chi}^2. \tag{C.9}$$

C.2 Curvature

The metric, affine connection and curvature tensors for the LTB metric (4.3) are:

1. **Metric:**

$$\begin{aligned} &g_{tt} = -1, \quad g_{rr} = \frac{A'^2(t,r)}{1-k(r)}, \\ &g_{\theta\theta} = A^2(t,r), \quad g_{\varphi\varphi} = A^2(t,r)\,\sin^2(\theta) \end{aligned} \tag{C.10}$$

2. **Affine connection:**

$$\begin{aligned} &\Gamma^t_{rr} = \frac{A'\,\dot{A}'}{1-k}, \quad \Gamma^t_{\theta\theta} = A\,\dot{A}, \quad \Gamma^t_{\varphi\varphi} = \Gamma^t_{\theta\theta}\sin^2(\theta), \quad \Gamma^r_{tr} = H_L \\ &\Gamma^r_{rr} = \frac{k'}{2(1-k)} + \frac{A''}{A'}, \quad \Gamma^r_{\theta\theta} = -\frac{A}{A'}(1-k\,r^2), \quad \Gamma^r_{\varphi\varphi} = \Gamma^r_{\theta\theta}\,\sin^2(\theta), \\ &\Gamma^\theta_{t\theta} = \Gamma^\varphi_{t\varphi} = H_T, \quad \Gamma^\theta_{r\theta} = \Gamma^\varphi_{r\varphi} = \frac{A'}{A}, \quad \Gamma^\theta_{\varphi\varphi} = -\sin\theta\,\cos\theta, \quad \Gamma^\varphi_{\theta\varphi} = \cot(\theta) \end{aligned} \tag{C.11}$$

3. **Riemann tensor:**

$$\begin{aligned} &R_{trtr} = \frac{A'\,\ddot{A}'}{1-k}, \quad R_{t\theta t\theta} = A\,\ddot{A}, \quad R_{t\varphi t\varphi} = A\,\ddot{A}\,\sin\theta \\ &R_{r\theta r\theta} = -\frac{A\,A'\,(k' + 2\dot{A}\dot{A}')}{2(1-k)}, \quad R_{r\varphi r\varphi} = R_{r\theta r\theta}\,\sin^2(\theta) \\ &R_{\theta\phi\theta\phi} = A^2(k + \dot{A}^2)\,\sin^2(\theta) \end{aligned} \tag{C.12}$$

4. **Ricci tensor:**

$$R_{tt} = \frac{2\ddot{A}}{A} + \frac{\ddot{A}'}{A'}, \quad R_{rr} = -\frac{A'(k' + 2\dot{A}\,\dot{A}' + A\,\ddot{A}')}{A(1-k)} \tag{C.13}$$

$$R_{\theta\theta} = -\left(k + \dot{A}^2 + A\,\ddot{A} + \frac{A\,(k' + 2\dot{A}\,\dot{A}')}{2A'}\right), \quad R_{\varphi\varphi} = R_{\theta\theta}\,\sin^2(\theta) \tag{C.14}$$

5. **Curvature scalar:**

$$R = \frac{2}{A^2A'}\left[A'\,(k + \dot{A}^2) + A(k' + 2\dot{A}\dot{A}' + 2A'\ddot{A}) + A^2\ddot{A}'\right] \tag{C.15}$$

6. **Einstein tensor:**

$$G_{tt} = -\frac{A'(k + \dot{A}^2) + A(k' + 2\dot{A}\dot{A}')}{A^2A'}, \tag{C.16}$$

$$G_{rr} = \frac{A'^2(k + \dot{A}^2 + 2A\ddot{A})}{A^2(1-k)}, \tag{C.17}$$

$$G_{\theta\theta} = \frac{A\left[k' + 2(\dot{A}\dot{A}' + A'\ddot{A} + A\ddot{A}')\right]}{2A'}, \quad G_{\varphi\varphi} = G_{\theta\theta}\,\sin^2(\theta) \tag{C.18}$$

Side note: If we do not assume (4.2) the component G_{tr} would be non-zero:

$$G_{tr} = \frac{2(X\dot{A}' - \dot{X}A')}{A\,X}. \tag{C.19}$$

Since in the absence of anisotropic stress (which is our case) $T_{tr} = 0$, Einstein's equations imply

$$\frac{\dot{X}}{X} = \frac{\dot{A}'}{A'} \;\Rightarrow\; X(t,r) = C(r)A'(t,r), \tag{C.20}$$

which motivates (4.2).

Bibliography

[1] E. Sanchez, A. Carnero, J. Garcia-Bellido, E. Gaztanaga, F. de Simoni, M. Crocce, A. Cabré, P. Fosalba, and D. Alonso, "Tracing The Sound Horizon Scale With Photometric Redshift Surveys," *Mon.Not.Roy.Astron.Soc.* **411** (2011) 277–288, arXiv:1006.3226 [astro-ph.CO].

[2] E. Sanchez, D. Alonso, F. Sanchez, J. Garcia-Bellido, and I. Sevilla, "Precise Measurement of the Radial Baryon Acoustic Oscillation Scales in Galaxy Redshift Surveys," arXiv:1210.6446 [astro-ph.CO].

[3] D. Adams, *The Hitchhiker's Guide to the Galaxy*. Pan MacMillan, 1979.

[4] **Planck Collaboration** Collaboration, P. Ade *et al.*, "Planck 2013 results. XVI. Cosmological parameters," arXiv:1303.5076 [astro-ph.CO].

[5] **Supernova Cosmology Project** Collaboration, S. Perlmutter *et al.*, "Measurements of Omega and Lambda from 42 high redshift supernovae," *Astrophys.J.* **517** (1999) 565–586, arXiv:astro-ph/9812133 [astro-ph].

[6] **Supernova Search Team** Collaboration, A. Riess *et al.*, "Observational evidence from supernovae for an accelerating universe and a cosmological constant," *Astron.J.* **116** (1998) 1009–1038, arXiv:astro-ph/9805201 [astro-ph].

[7] D. Alonso, J. Garcia-Bellido, T. Haugbolle, and J. Vicente, "Large scale structure simulations of inhomogeneous LTB void models," *Phys.Rev.* **D82** (2010) 123530, arXiv:1010.3453 [astro-ph.CO].

[8] D. Alonso, J. Garcia-Bellido, T. Haugboelle, and A. Knebe, "Halo abundances and shear in void models," *Phys.Dark Univ.* **1** (2012) 24–31, arXiv:1204.3532 [astro-ph.CO].

[9] E. Hubble, "A relation between distance and radial velocity among extra-galactic nebulae," *Proc.Nat.Acad.Sci.* **15** (1929) 168–173.

[10] A. Guth, "The Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems," *Phys.Rev.* **D23** (1981) 347–356.

[11] A. Linde, *Inflation and quantum cosmology*. Academic Press, 1990.

[12] S. Weinberg, *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*. John Wiley & Sons, Inc., 1972.

[13] A. Friedman, "On the Curvature of space," *Z.Phys.* **10** (1922) 377–386.

[14] G. Bertone, D. Hooper, and J. Silk, "Particle dark matter: Evidence, candidates and constraints," *Phys.Rept.* **405** (2005) 279–390, arXiv:hep-ph/0404175 [hep-ph].

[15] M. Hobson, G. Efstathiou, and A. Lasenby, *General Relativity: An Introduction for Physicists*. Cambridge University Press, 2006.

[16] J. Garcia-Bellido, "Cosmology and astrophysics," arXiv:astro-ph/0502139 [astro-ph].

[17] G. Ellis, "Relativistic cosmology," *Gen.Rel.Grav.* **41** (2009) 581–660.

[18] E. Kolb and M. Turner, *The Early Universe*. Westview Press, 1994.

[19] V. Mukhanov, *Physical Foundations of Cosmology*. Cambridge University Press, 2005.

[20] R. Jimenez, P. Thejll, U. Jorgensen, J. MacDonald, and B. Pagel, "Ages of globular clusters: a new approach," *Mon.Not.Roy.Astron.Soc.* **282** (1996) 926–942, arXiv:astro-ph/9602132 [astro-ph].

[21] G. Hinshaw, D. Larson, E. Komatsu, D. Spergel, C. Bennett, *et al.*, "Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results," arXiv:1212.5226 [astro-ph.CO].

[22] M. Hamuy, M. Phillips, R. Schommer, N. Suntzeff, J. Maza, *et al.*, "The Absolute luminosities of the Calan/Tololo type IA supernovae," *Astron.J.* **112** (1996) 2391, arXiv:astro-ph/9609059 [astro-ph].

[23] **SNLS Collaboration** Collaboration, P. Astier *et al.*, "The Supernova legacy survey: Measurement of omega(m), omega(lambda) and W from the first year data set," *Astron.Astrophys.* **447** (2006) 31–48, arXiv:astro-ph/0510447 [astro-ph].

[24] **Supernova Cosmology Project** Collaboration, M. Kowalski *et al.*, "Improved Cosmological Constraints from New, Old and Combined Supernova Datasets," *Astrophys.J.* **686** (2008) 749–778, arXiv:0804.4142 [astro-ph].

[25] **SDSS Collaboration** Collaboration, W. Percival *et al.*, "Baryon Acoustic Oscillations in the Sloan Digital Sky Survey Data Release 7 Galaxy Sample," *Mon.Not.Roy.Astron.Soc.* **401** (2010) 2148–2168, arXiv:0907.1660 [astro-ph.CO].

[26] N. Padmanabhan, X. Xu, D. Eisenstein, R. Scalzo, A. Cuesta, *et al.*, "A 2Methods and Application to the Sloan Digital Sky Survey," arXiv:1202.0090 [astro-ph.CO].

[27] C. Blake, C. Kazin, F. Beutler, T. Davis, D. Parkinson, *et al.*, "The WiggleZ Dark Energy Survey: mapping the distance-redshift relation with baryon acoustic oscillations," *Mon.Not.Roy.Astron.Soc.* **418** (2011) 1707–1724, arXiv:1108.2635 [astro-ph.CO].

[28] F. Beutler, C. Blake, M. Colless, D. Jones, L. Staveley-Smith, *et al.*, "The 6dF Galaxy Survey: Baryon Acoustic Oscillations and the Local Hubble Constant," *Mon.Not.Roy.Astron.Soc.* **416** (2011) 3017–3032, arXiv:1106.3366 [astro-ph.CO].

[29] A. Carnero, E. Sánchez, M. Crocce, A. Cabré, and E. Gaztañaga, "Clustering of Photometric Luminous Red Galaxies II: Cosmological Implications from the Baryon Acoustic Scale," *Mon.Not.Roy.Astron.Soc.* **419** (2012) 1689–1694, arXiv:1104.5426 [astro-ph.CO].

[30] L. Anderson, E. Aubourg, S. Bailey, D. Bizyaev, M. Blanton, *et al.*, "The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Baryon Acoustic Oscillations in the Data Release 9 Spectroscopic Galaxy Sample," *Mon.Not.Roy.Astron.Soc.* **428** (2013) 1036–1054, arXiv:1203.6594 [astro-ph.CO].

[31] R. Durrer, *The Cosmic Microwave Background*. Cambridge University Press, 2008.

[32] **WMAP Collaboration** Collaboration, E. Komatsu *et al.*, "Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation," *Astrophys.J.Suppl.* **192** (2011) 18, arXiv:1001.4538 [astro-ph.CO].

[33] A. Sakharov, "Vacuum quantum fluctuations in curved space and the theory of gravitation," *Sov.Phys.Dokl.* **12** (1968) 1040–1041.

[34] J. Martin, "Everything You Always Wanted To Know About The Cosmological Constant Problem (But Were Afraid To Ask)," *Comptes Rendus Physique* **13** (2012) 566–665, arXiv:1205.3365 [astro-ph.CO].

[35] L. Susskind, "The Anthropic landscape of string theory," arXiv:hep-th/0302219 [hep-th].

[36] S. Weinberg, "Anthropic Bound on the Cosmological Constant," *Phys.Rev.Lett.* **59** (1987) 2607.

[37] E. Copeland, M. Sami, and S. Tsujikawa, "Dynamics of dark energy," *Int.J.Mod.Phys.* **D15** (2006) 1753–1936, arXiv:hep-th/0603057 [hep-th].

[38] S. Tsujikawa, "Dark energy: investigation and modeling," arXiv:1004.1493 [astro-ph.CO].

[39] L. Amendola and S. Tsujikawa, *Dark Energy: Theory and Observations*. Cambridge University Press, 2010.

[40] R. Caldwell, R. Dave, and P. Steinhardt, "Cosmological imprint of an energy component with general equation of state," *Phys.Rev.Lett.* **80** (1998) 1582–1585, arXiv:astro-ph/9708069 [astro-ph].

[41] E. Linder, "Exploring the expansion history of the universe," *Phys.Rev.Lett.* **90** (2003) 091301, arXiv:astro-ph/0208512 [astro-ph].

[42] M. Chevallier and D. Polarski, "Accelerating universes with scaling dark matter," *Int.J.Mod.Phys.* **D10** (2001) 213–224, arXiv:gr-qc/0009008 [gr-qc].

[43] T. Chiba, T. Okabe, and M. Yamaguchi, "Kinetically driven quintessence," *Phys.Rev.* **D62** (2000) 023511, arXiv:astro-ph/9912463 [astro-ph].

[44] J. Khoury and A. Weltman, "Chameleon cosmology," *Phys.Rev.* **D69** (2004) 044026, arXiv:astro-ph/0309411 [astro-ph].

[45] **CMS Collaboration** Collaboration, "A new boson with a mass of 125-GeV observed with the CMS experiment at the Large Hadron Collider," *Science* **338** (2012) 1569–1575.

[46] **ATLAS Collaboration** Collaboration, G. Aad *et al.*, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC," *Phys.Lett.* **B716** (2012) 1–29, arXiv:1207.7214 [hep-ex].

[47] J. J. Beltran and A. Maroto, "A cosmic vector for dark energy," *Phys.Rev.* **D78** (2008) 063005, arXiv:0801.1486 [astro-ph].

[48] A. Kamenshchik, U. Moschella, and V. Pasquier, "An Alternative to quintessence," *Phys.Lett.* **B511** (2001) 265–268, arXiv:gr-qc/0103004 [gr-qc].

[49] V. Faraoni, "Solar System experiments do not yet veto modified gravity models," *Phys.Rev.* **D74** (2006) 023529, arXiv:gr-qc/0607016 [gr-qc].

[50] A. Starobinsky, "A New Type of Isotropic Cosmological Models Without Singularity," *Phys.Lett.* **B91** (1980) 99–102.

[51] D. Vollick, "1/R Curvature corrections as the source of the cosmological acceleration," *Phys.Rev.* **D68** (2003) 063510, arXiv:astro-ph/0306630 [astro-ph].

[52] T. Clifton, P. Ferreira, A. Padilla, and C. Skordis, "Modified Gravity and Cosmology," *Phys.Rept.* **513** (2012) 1–189, arXiv:1106.2476 [astro-ph.CO].

[53] R. Durrer, "What do we really know about Dark Energy?," *Phil.Trans.Roy.Soc.Lond.* **A369** (2011) 5102–5114, arXiv:1103.5331 [astro-ph.CO].

[54] S. Rasanen, "Dark energy from backreaction," *JCAP* **0402** (2004) 003, arXiv:astro-ph/0311257 [astro-ph].

[55] E. Kolb, S. Matarrese, A. Notari, and A. Riotto, "The Effect of inhomogeneities on the expansion rate of the universe," *Phys.Rev.* **D71** (2005) 023524, arXiv:hep-ph/0409038 [hep-ph].

[56] E. Kolb, S. Matarrese, and A. Riotto, "On cosmic acceleration without dark energy," *New J.Phys.* **8** (2006) 322, arXiv:astro-ph/0506534 [astro-ph].

[57] C. Hirata and U. Seljak, "Can superhorizon cosmological perturbations explain the acceleration of the Universe?," *Phys.Rev.* **D72** (2005) 083501, arXiv:astro-ph/0503582 [astro-ph].

[58] P. Martineau and R. Brandenberger, "The Effects of gravitational back-reaction on cosmological perturbations," *Phys.Rev.* **D72** (2005) 023507, arXiv:astro-ph/0505236 [astro-ph].

[59] M. Seikel and D. Schwarz, "Probing Backreaction Effects with Supernova Data," arXiv:0912.2308 [astro-ph.CO].

[60] K. Tomita, "A local void and the accelerating universe," *Mon.Not.Roy.Astron.Soc.* **326** (2001) 287, arXiv:astro-ph/0011484 [astro-ph].

[61] M. Celerier, "Do we really see a cosmological constant in the supernovae data?," *Astron.Astrophys.* **353** (2000) 63–71, arXiv:astro-ph/9907206 [astro-ph].

[62] A. Romano and M. Sasaki, "Spatial averaging and apparent acceleration in inhomogeneous spaces," *Gen.Rel.Grav.* **44** (2012) 353–365, arXiv:0905.3342 [astro-ph.CO].

[63] K. Enqvist, "Lemaitre-Tolman-Bondi model and accelerating expansion," *Gen.Rel.Grav.* **40** (2008) 451–466, arXiv:0709.2044 [astro-ph].

[64] J. Garcia-Bellido and T. Haugboelle, "Confronting Lemaitre-Tolman-Bondi models with Observational Cosmology," *JCAP* **0804** (2008) 003, arXiv:0802.1523 [astro-ph].

[65] J. Zibin, A. Moss, and D. Scott, "Can we avoid dark energy?," *Phys.Rev.Lett.* **101** (2008) 251303, arXiv:0809.3761 [astro-ph].

[66] C. Clarkson, "Establishing homogeneity of the universe in the shadow of dark energy," *Comptes Rendus Physique* **13** (2012) 682–718, arXiv:1204.5505 [astro-ph.CO].

[67] A. Liddle and D. Lyth, *Cosmological Inflation and Large-Scale Structure*. Cambridge University Press, 2000.

[68] S. Tsujikawa, "Introductory review of cosmic inflation," arXiv:hep-ph/0304257 [hep-ph].

[69] J. H. Jeans, "The Stability of a Spherical Nebula," *Philosophical Transactions of the Royal Society of London* **Series A 199** 1.

[70] J. Garcia-Bellido and D. Wands, "Metric perturbations in two field inflation," *Phys.Rev.* **D53** (1996) 5437–5445, arXiv:astro-ph/9511029 [astro-ph].

[71] S. Carroll, W. Press, and E. Turner, "The Cosmological constant," *Ann.Rev.Astron.Astrophys.* **30** (1992) 499–542.

[72] A. B. Belloso, J. Garcia-Bellido, and D. Sapone, "A parametrization of the growth index of matter perturbations in various Dark Energy models and observational prospects using a Euclid-like survey," *JCAP* **1110** (2011) 010, arXiv:1105.4825 [astro-ph.CO].

[73] P. Meszaros, "The behaviour of point masses in an expanding cosmological substratum," *Astron.Astrophys.* **37** (1974) 225–228.

[74] S. Lee and K. Ng, "Growth index with the exact analytic solution of sub-horizon scale linear perturbation for dark energy models with constant equation of state," *Phys.Lett.* **B688** (2010) 1–3, arXiv:0906.1643 [astro-ph.CO].

[75] J. Bardeen, J. Bond, N. Kaiser, and A. Szalay, "The Statistics of Peaks of Gaussian Random Fields," *Astrophys.J.* **304** (1986) 15–61.

[76] H. Mo and S. White, "An Analytic model for the spatial clustering of dark matter halos," *Mon.Not.Roy.Astron.Soc.* **282** (1996) 347, arXiv:astro-ph/9512127 [astro-ph].

[77] A. Hamilton, "Linear redshift distortions: A Review," arXiv:astro-ph/9708102 [astro-ph].

[78] N. Kaiser, "Clustering in real space and in redshift space," *Mon.Not.Roy.Astron.Soc.* **227** (1987) 1–27.

[79] M. Davis and P. Peebles, "A Survey of galaxy redshifts. 5. The Two point position and velocity correlations," *Astrophys.J.* **267** (1982) 465–482.

[80] J. Fry, "The Galaxy correlation hierarchy in perturbation theory," *Astrophys.J.* **279** (1984) 499–510.

[81] J. Peacock and S. Dodds, "Reconstructing the linear power spectrum of cosmological mass fluctuations," *Mon.Not.Roy.Astron.Soc.* **267** (1994) 1020–1034, arXiv:astro-ph/9311057 [astro-ph].

[82] R. Scoccimarro, "Redshift-space distortions, pairwise velocities and nonlinearities," *Phys.Rev.* **D70** (2004) 083007, arXiv:astro-ph/0407214 [astro-ph].

[83] L. Guzzo, M. Pierleoni, B. Meneux, E. Branchini, O. Fevre, *et al.*, "A test of the nature of cosmic acceleration using galaxy redshift distortions," *Nature* **451** (2008) 541–545, arXiv:0802.1944 [astro-ph].

[84] L. Samushia, W. Percival, and A. Raccanelli, "Interpreting large-scale redshift-space distortion measurements," *Mon.Not.Roy.Astron.Soc.* **420** (2012) 2102–2119, arXiv:1102.1014 [astro-ph.CO].

[85] C. Chuang, F. Prada, A. Cuesta, D. Eisenstein, E. Kazin, *et al.*, "The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: single-probe measurements and the strong power of normalized growth rate on constraining dark energy," arXiv:1303.4486 [astro-ph.CO].

[86] H. Kodama and M. Sasaki, "Cosmological Perturbation Theory," *Prog.Theor.Phys.Suppl.* **78** (1984) 1–166.

[87] S. Dodelson, *General Relativity: An Introduction for Physicists*. Academic Press, 2003.

[88] A. Lewis, A. Challinor, and A. Lasenby, "Efficient computation of CMB anisotropies in closed FRW models," *Astrophys.J.* **538** (2000) 473–476, arXiv:astro-ph/9911177 [astro-ph].

[89] J. Lesgourgues, "The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview," arXiv:1104.2932 [astro-ph.IM].

[90] J. Silk, "Cosmic black body radiation and galaxy formation," *Astrophys.J.* **151** (1968) 459–471.

[91] D. Eisenstein and W. Hu, "Baryonic features in the matter transfer function," *Astrophys.J.* **496** (1998) 605, arXiv:astro-ph/9709112 [astro-ph].

[92] F. Bernardeau, S. Colombi, E. Gaztanaga, and R. Scoccimarro, "Large scale structure of the universe and cosmological perturbation theory," *Phys.Rept.* **367** (2002) 1–248, arXiv:astro-ph/0112551 [astro-ph].

[93] M. Goroff, B. Grinstein, S. Rey, and M. Wise, "Coupling of Modes of Cosmological Mass Density Fluctuations," *Astrophys.J.* **311** (1986) 6–14.

[94] E. Bertschinger and B. Jain, "Gravitational instability of cold matter," *Astrophys.J.* **431** (1994) 486, arXiv:astro-ph/9307033 [astro-ph].

[95] F. Bouchet, "Introductory overview of Eulerian and Lagrangian perturbation theories," arXiv:astro-ph/9603013 [astro-ph].

[96] T. Buchert, "Lagrangian perturbation approach to the formation of large scale structure," arXiv:astro-ph/9509005 [astro-ph].

[97] Y. Zeldovich, "Gravitational instability: An Approximate theory for large density perturbations," *Astron.Astrophys.* **5** (1970) 84.

[98] D. Pogosian, J. Bond, L. Kofman, and J. Wadsley, "Cosmic web: Origin and observables," arXiv:astro-ph/9810072 [astro-ph].

[99] M. Crocce, S. Pueblas, and R. Scoccimarro, "Transients from Initial Conditions in Cosmological Simulations," *Mon.Not.Roy.Astron.Soc.* **373** (2006) 369–381, arXiv:astro-ph/0606505 [astro-ph].

[100] R. Mandelbaum, U. Seljak, G. Kauffmann, C. Hirata, and J. Brinkmann, "Galaxy halo masses and satellite fractions from galaxy-galaxy lensing in the sdss: stellar mass, luminosity, morphology, and environment dependencies," *Mon.Not.Roy.Astron.Soc.* **368** (2006) 715, arXiv:astro-ph/0511164 [astro-ph].

[101] J. Peacock and R. Smith, "Halo occupation numbers and galaxy bias," *Mon.Not.Roy.Astron.Soc.* **318** (2000) 1144, arXiv:astro-ph/0005010 [astro-ph].

[102] C. Ma and J. Fry, "Deriving the nonlinear cosmological power spectrum and bispectrum from analytic dark matter halo profiles and mass functions," *Astrophys.J.* **543** (2000) 503.

[103] U. Seljak, "Analytic model for galaxy and dark matter clustering," *Mon.Not.Roy.Astron.Soc.* **318** (2000) 203, arXiv:astro-ph/0001493 [astro-ph].

[104] W. Press and P. Schechter, "Formation of galaxies and clusters of galaxies by selfsimilar gravitational condensation," *Astrophys.J.* **187** (1974) 425–438.

[105] H. Mo, F. van den Bosch, and S. White, *Galaxy Formation and Evolution*. Cambridge University Press, 2010.

[106] N. Weinberg and M. Kamionkowski, "Constraining dark energy from the abundance of weak gravitational lenses," *Mon.Not.Roy.Astron.Soc.* **341** (2003) 251, arXiv:astro-ph/0210134 [astro-ph].

[107] A. Jenkins, C. Frenk, S. White, J. Colberg, S. Cole, *et al.*, "The Mass function of dark matter halos," *Mon.Not.Roy.Astron.Soc.* **321** (2001) 372, arXiv:astro-ph/0005260 [astro-ph].

[108] M. Warren, K. Abazajian, D. Holz, and L. Teodoro, "Precision determination of the mass function of dark matter halos," *Astrophys.J.* **646** (2006) 881–885, arXiv:astro-ph/0506395 [astro-ph].

[109] J. Tinker, A. Kravtsov, A. Klypin, K. Abazajian, M. Warren, *et al.*, "Toward a halo mass function for precision cosmology: The Limits of universality," *Astrophys.J.* **688** (2008) 709–728, arXiv:0803.2706 [astro-ph].

[110] R. Sheth, H. Mo, and G. Tormen, "Ellipsoidal collapse and an improved model for the number and spatial distribution of dark matter haloes," *Mon.Not.Roy.Astron.Soc.* **323** (2001) 1, arXiv:astro-ph/9907024 [astro-ph].

[111] J. Bond, S. Cole, G. Efstathiou, and N. Kaiser, "Excursion set mass functions for hierarchical Gaussian fluctuations," *Astrophys.J.* **379** (1991) 440.

[112] J. Peacock and A. Heavens, "Alternatives to the Press-Schechter cosmological mass function," *Mon.Not.Roy.Astron.Soc.* **243** (1990) 133–143.

[113] M. Musso and R. Sheth, "One step beyond: The excursion set approach with correlated steps," *Mon.Not.Roy.Astron.Soc.* **423** (2012) L102–L106, arXiv:1201.3876 [astro-ph.CO].

[114] C. Lacey and S. Cole, "Merger rates in hierarchical models of galaxy formation," *Mon.Not.Roy.Astron.Soc.* **262** (1993) 627–649.

[115] C. Lacey and S. Cole, "Merger rates in hierarchical models of galaxy formation. 2. Comparison with N body simulations," *Mon.Not.Roy.Astron.Soc.* **271** (1994) 676, arXiv:astro-ph/9402069 [astro-ph].

[116] S. Cole, J. Helly, C. Frenk, and H. Parkinson, "The statistical properties of LCDM halo formation," *Mon.Not.Roy.Astron.Soc.* (2007) , arXiv:0708.1376 [astro-ph].

[117] M. Musso, A. Paranjape, and R. Sheth, "Scale dependent halo bias in the excursion set approach," *Mon.Not.Roy.Astron.Soc.* **427** (2012) 3145–3158, arXiv:1205.3401 [astro-ph.CO].

[118] J. Fry and E. Gaztañaga, "Biasing and hierarchical statistics in large scale structure," *Astrophys.J.* **413** (1993) 447–452, arXiv:astro-ph/9302009 [astro-ph].

[119] T. Giannantonio, C. Porciani, J. Carron, A. Amara, and A. Pillepich, "Constraining primordial non-Gaussianity with future galaxy surveys," *Mon.Not.Roy.Astron.Soc.* **422** (2012) 2854–2877, arXiv:1109.0958 [astro-ph.CO].

[120] A. Dekel and O. Lahav, "Stochastic nonlinear galaxy biasing," *Astrophys.J.* **520** (1999) 24–34, arXiv:astro-ph/9806193 [astro-ph].

[121] J. Navarro, C. Frenk, and S. White, "A Universal density profile from hierarchical clustering," *Astrophys.J.* **490** (1997) 493–508, arXiv:astro-ph/9611107 [astro-ph].

[122] J. Einasto, "Kinematics and dynamics of stellar systems," *Trudy Inst. Astrofiz. Alma-Ata* **51** 87.

[123] J. Navarro, E. Hayashi, C. Power, A. Jenkins, C. Frenk, *et al.*, "The Inner structure of Lambda-CDM halos 3: Universality and asymptotic slopes," *Mon.Not.Roy.Astron.Soc.* **349** (2004) 1039, arXiv:astro-ph/0311231 [astro-ph].

[124] D. Figueroa, E. Sefusatti, A. Riotto, and F. Vernizzi, "The Effect of Local non-Gaussianity on the Matter Bispectrum at Small Scales," *JCAP* **1208** (2012) 036, arXiv:1205.2015 [astro-ph.CO].

[125] **Virgo Consortium** Collaboration, R. Smith *et al.*, "Stable clustering, the halo model and nonlinear cosmological power spectra," *Mon.Not.Roy.Astron.Soc.* **341** (2003) 1311, arXiv:astro-ph/0207664 [astro-ph].

[126] V. Springel, "The Cosmological simulation code GADGET-2," *Mon.Not.Roy.Astron.Soc.* **364** (2005) 1105–1134, arXiv:astro-ph/0505010 [astro-ph].

[127] W. Dehnen, "Towards optimal softening in 3-D n-body codes: I. minimizing the force error," *Mon.Not.Roy.Astron.Soc.* **324** (2001) 273, arXiv:astro-ph/0011568 [astro-ph].

[128] R. Hockney and J. Eastwood, *Computer Simulation Using Particles*. Taylor & Francis, 1988.

[129] A. Klypin, "Numerical simulations in cosmology I: methods," arXiv:astro-ph/0005502 [astro-ph].

[130] M. Kuhlen, M. Vogelsberger, and R. Angulo, "Numerical Simulations of the Dark Universe: State of the Art and the Next Decade," *Phys.Dark Univ.* **1** (2012) 50–93, arXiv:1209.5745 [astro-ph.CO].

[131] D. Eisenstein and C. Bennett, "Cosmic sound waves rule," *Phys.Today* **61N4** (2008) 44–50.

[132] D. Weinberg, M. Mortonson, D. Eisenstein, C. Hirata, A. Riess, *et al.*, "Observational Probes of Cosmic Acceleration," arXiv:1201.2434 [astro-ph.CO].

[133] **SDSS Collaboration** Collaboration, D. Eisenstein *et al.*, "Detection of the baryon acoustic peak in the large-scale correlation function of SDSS luminous red galaxies," *Astrophys.J.* **633** (2005) 560–574, arXiv:astro-ph/0501171 [astro-ph].

[134] **2dFGRS Collaboration** Collaboration, S. Cole *et al.*, "The 2dF Galaxy Redshift Survey: Power-spectrum analysis of the final dataset and cosmological implications," *Mon.Not.Roy.Astron.Soc.* **362** (2005) 505–534, arXiv:astro-ph/0501174 [astro-ph].

[135] **2DFGRS Collaboration** Collaboration, M. Colless *et al.*, "The 2dF Galaxy Redshift Survey: Spectra and redshifts," *Mon.Not.Roy.Astron.Soc.* **328** (2001) 1039, arXiv:astro-ph/0106498 [astro-ph].

[136] D. Jones, W. Saunders, M. Colless, M. Read, Q. Parker, *et al.*, "The 6dF Galaxy Survey: Samples, observational techniques and the first data release," *Mon.Not.Roy.Astron.Soc.* **355** (2004) 747–763, arXiv:astro-ph/0403501 [astro-ph].

[137] M. Drinkwater, R. Jurek, C. Blake, D. Woods, K. Pimbblet, *et al.*, "The WiggleZ Dark Energy Survey: Survey Design and First Data Release," *Mon.Not.Roy.Astron.Soc.* **401** (2010) 1429–1452, arXiv:0911.4246 [astro-ph.CO].

[138] **SDSS Collaboration** Collaboration, C. Ahn *et al.*, "The Ninth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the SDSS-III Baryon Oscillation Spectroscopic Survey," *Astrophys.J.Suppl.* **203** (2012) 21, arXiv:1207.7137 [astro-ph.IM].

[139] M. Bolzonella, J. Miralles, and R. Pello, "Photometric redshifts based on standard SED fitting procedures," *Astron.Astrophys.* **363** (2000) 476–492, arXiv:astro-ph/0003380 [astro-ph].

[140] N. Benitez, "Bayesian photometric redshift estimation," *Astrophys.J.* **536** (2000) 571–583, arXiv:astro-ph/9811189 [astro-ph].

[141] A. Firth, O. Lahav, and R. Somerville, "Estimating photometric redshifts with artificial neural networks," *Mon.Not.Roy.Astron.Soc.* **339** (2003) 1195, arXiv:astro-ph/0203250 [astro-ph].

[142] C. Blake and S. Bridle, "Cosmology with photometric redshift surveys," *Mon.Not.Roy.Astron.Soc.* **363** (2005) 1329–1348, arXiv:astro-ph/0411713 [astro-ph].

[143] **Dark Energy Survey Collaboration** Collaboration, T. Abbott *et al.*, "The dark energy survey," arXiv:astro-ph/0510346 [astro-ph].

[144] N. Kaiser, H. Aussel, H. Boesgaard, K. Chambers, J. Heasley, *et al.*, "Pan-STARRS: A Large Synoptic Survey Telescope Array," *Proc.SPIE Int.Soc.Opt.Eng.* **4836** (2002) 154–164.

[145] **LSST Collaboration** Collaboration, J. Tyson, "Large synoptic survey telescope: Overview," *Proc.SPIE Int.Soc.Opt.Eng.* **4836** (2002) 10–20, arXiv:astro-ph/0302102 [astro-ph].

[146] E. Gaztanaga, M. Eriksen, M. Crocce, F. Castander, P. Fosalba, *et al.*, "Cross-Correlation of spectroscopic and photometric galaxy surveys: cosmology from lensing and redshift distortions," arXiv:1109.4852 [astro-ph.CO].

[147] N. Benitez, E. Gaztanaga, R. Miquel, F. Castander, M. Moles, *et al.*, "Measuring Baryon Acoustic Oscillations along the line of sight with photometric redshifs: the PAU survey," *Astrophys.J.* **691** (2009) 241–260, arXiv:0807.0535 [astro-ph].

[148] M. Crocce and R. Scoccimarro, "Nonlinear Evolution of Baryon Acoustic Oscillations," *Phys.Rev.* **D77** (2008) 023533, arXiv:0704.2783 [astro-ph].

[149] C. Alcock and B. Paczynski, "An evolution free test for non-zero cosmological constant," *Nature* **281** (1979) 358–359.

[150] V. Martínez and E. Saar, *Statistics of the galaxy distribution.* CRC Press, 2002.

[151] M. Manera, R. Scoccimarro, W. Percival, L. Samushia, C. McBride, *et al.*, "The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: a large sample of mock galaxy catalogues," arXiv:1203.6609 [astro-ph.CO].

[152] S. de la Torre and J. Peacock, "Reconstructing the distribution of haloes and mock galaxies below the resolution limit in cosmological simulations," arXiv:1212.3615 [astro-ph.CO].

[153] P. Coles and B. Jones, "A Lognormal model for the cosmological mass distribution," *Mon.Not.Roy.Astron.Soc.* **248** (1991) 1–13.

[154] W. Percival, L. Verde, and J. Peacock, "Fourier analysis of luminosity-dependent galaxy clustering," *Mon.Not.Roy.Astron.Soc.* **347** (2004) 645, arXiv:astro-ph/0306511 [astro-ph].

[155] A. Berlind and D. Weinberg, "The Halo occupation distribution: Towards an empirical determination of the relation between galaxies and mass," *Astrophys.J.* **575** (2002) 587–616, arXiv:astro-ph/0109001 [astro-ph].

[156] R. Scoccimarro, R. Sheth, L. Hui, and B. Jain, "How many galaxies fit in a halo? Constraints on galaxy formation efficiency from spatial clustering," *Astrophys.J.* **546** (2001) 20–34, arXiv:astro-ph/0006319 [astro-ph].

[157] D. Eisenstein, H. Seo, and . White, M.J., "On the Robustness of the Acoustic Scale in the Low-Redshift Clustering of Matter," *Astrophys.J.* **664** (2007) 660–674, arXiv:astro-ph/0604361 [astro-ph].

[158] N. Padmanabhan and M. White, "Calibrating the Baryon Oscillation Ruler for Matter and Halos," *Phys.Rev.* **D80** (2009) 063508, arXiv:0906.1198 [astro-ph.CO].

[159] H. Seo, J. Eckel, D. Eisenstein, K. Mehta, M. Metchnik, *et al.*, "High-precision predictions for the acoustic scale in the non-linear regime," *Astrophys.J.* **720** (2010) 1650–1667, arXiv:0910.5005 [astro-ph.CO].

[160] D. Eisenstein, H. Seo, E. Sirko, and D. Spergel, "Improving Cosmological Distance Measurements by Reconstruction of the Baryon Acoustic Peak," *Astrophys.J.* **664** (2007) 675–679, arXiv:astro-ph/0604362 [astro-ph].

[161] K. Mehta, H. Seo, J. Eckel, D. Eisenstein, M. Metchnik, *et al.*, "Galaxy Bias and its Effects on the Baryon Acoustic Oscillations Measurements," *Astrophys.J.* **734** (2011) 94, arXiv:1104.1178 [astro-ph.CO].

[162] O. Ilbert, S. Arnouts, H. McCracken, M. Bolzonella, E. Bertin, *et al.*, "Accurate photometric redshifts for the cfht legacy survey calibrated using the vimos vlt deep survey," *Astron.Astrophys.* **457** (2006) 841–856, arXiv:astro-ph/0603217 [astro-ph].

[163] M. Crocce, A. Cabre, and E. Gaztanaga, "Modeling the angular correlation function and its full covariance in Photometric Galaxy Surveys," *Mon.Not.Roy.Astron.Soc.* **414** (2011) 329–349, arXiv:1004.4640 [astro-ph.CO].

[164] H. Seo and D. Eisenstein, "Improved forecasts for the baryon acoustic oscillations and cosmological distance scale," *Astrophys.J.* **665** (2007) 14–24, arXiv:astro-ph/0701079 [astro-ph].

[165] A. Sanchez, C. Baugh, and R. Angulo, "What is the best way to measure baryonic acoustic oscillations?," *Mon.Not.Roy.Astron.Soc.* **390** (2008) 1470–1490, arXiv:0804.0233 [astro-ph].

[166] H. Seo, E. Siegel, D. Eisenstein, and M. White, "Non-linear structure formation and the acoustic scale," *Astrophys.J.* **686** (2008) 13–24, arXiv:0805.0117 [astro-ph].

[167] K. Nock, W. Percival, and A. Ross, "The effect of redshift-space distortions on projected 2-pt clustering measurements," *Mon.Not.Roy.Astron.Soc.* **407** (2010) 520, arXiv:1003.0896 [astro-ph.CO].

[168] M. LoVerde, L. Hui, and E. Gaztanaga, "Lensing corrections to features in the angular two-point correlation function and power spectrum," *Phys.Rev.* **D77** (2008) 023512, arXiv:0708.0031 [astro-ph].

[169] F. Simpson, J. Peacock, and P. Simon, "Locating the Baryon Acoustic Peak," *Phys.Rev.* **D79** (2009) 063508, arXiv:0901.3085 [astro-ph.CO].

[170] R. Smith, R. Scoccimarro, and R. Sheth, "Eppur Si Muove: On The Motion of the Acoustic Peak in the Correlation Function," *Phys.Rev.* **D77** (2008) 043525, arXiv:astro-ph/0703620 [ASTRO-PH].

[171] E. Gaztanaga, A. Cabre, and L. Hui, "Clustering of Luminous Red Galaxies IV: Baryon Acoustic Peak in the Line-of-Sight Direction and a Direct Measurement of H(z)," *Mon.Not.Roy.Astron.Soc.* **399** (2009) 1663–1680, arXiv:0807.3551 [astro-ph].

[172] **DES** Collaboration, T. Abbot *et al.*, "First SN Discoveries from the Dark Energy Survey," *A.Tel.* **4668** .

[173] P. Fosalba, E. Gaztanaga, F. Castander, and M. Manera, "The onion universe: all sky light-cone simulations in shells," *Mon.Not.Roy.Astron.Soc.* **391** (2007) 435, arXiv:0711.1540 [astro-ph].

[174] M. Crocce, P. Fosalba, F. Castander, and E. Gaztanaga, "Simulating the Universe with MICE: The abundance of massive clusters," *Mon.Not.Roy.Astron.Soc.* **403** (2010) 1353–1367, arXiv:0907.0019 [astro-ph.CO].

[175] M. Banerji, F. Abdalla, O. Lahav, and H. Lin, "Photometric Redshifts for the Dark Energy Survey and VISTA and Implications for Large Scale Structure," *Mon.Not.Roy.Astron.Soc.* **386** (2008) 1219–1233, arXiv:0711.1059 [astro-ph].

[176] S. Landy and A. Szalay, "Bias and variance of angular correlation functions," *Astrophys.J.* **412** (1993) 64.

[177] J. Benjamin, L. Van Waerbeke, B. Menard, and M. Kilbinger, "Photometric redshifts: estimating their contamination and distribution using clustering information," *Mon.Not.Roy.Astron.Soc.* **408** (2010) 1168–1180, arXiv:1002.2266 [astro-ph.CO].

[178] J. Cresswell and W. Percival, "Scale Dependent Galaxy Bias in the SDSS as a function of Luminosity and Colour," *Mon.Not.Roy.Astron.Soc.* **392** (2009) 682–690, arXiv:0808.1101 [astro-ph].

[179] E. Gaztanaga, R. Miquel, and E. Sanchez, "First Cosmological Constraints on Dark Energy from the Radial Baryon Acoustic Scale," *Phys.Rev.Lett.* **103** (2009) 091302, arXiv:0808.1921 [astro-ph].

[180] **SDSS Collaboration** Collaboration, E. Kazin *et al.*, "The Baryonic Acoustic Feature and Large-Scale Clustering in the SDSS LRG Sample," *Astrophys.J.* **710** (2010) 1444–1461, arXiv:0908.2598 [astro-ph.CO].

[181] F. Abdalla, J. Annis, D. Bacon, S. Bridle, F. Castander, *et al.*, "The Dark Energy Spectrometer (DESpec): A Multi-Fiber Spectroscopic Upgrade of the Dark Energy Camera and Survey for the Blanco Telescope," arXiv:1209.2451 [astro-ph.CO].

[182] **BigBoss Experiment** Collaboration, D. Schelgel *et al.*, "The BigBOSS Experiment," arXiv:1106.1706 [astro-ph.IM].

[183] R. Laureijs, J. Amiaux, S. Arduini, J.-L. Augueres, J. Brinchmann, *et al.*, "Euclid Definition Study Report," arXiv:1110.3193 [astro-ph.CO].

[184] J. Asorey, M. Crocce, E. Gaztanaga, and A. Lewis, "Recovering 3D clustering information with angular correlations," arXiv:1207.6487 [astro-ph.CO].

[185] C. Bonvin and R. Durrer, "What galaxy surveys really measure," *Phys.Rev.* **D84** (2011) 063505, arXiv:1105.5280 [astro-ph.CO].

[186] F. Montanari and R. Durrer, "A new method for the Alcock-Paczynski test," *Phys.Rev.* **D86** (2012) 063503, arXiv:1206.3545 [astro-ph.CO].

[187] **WMAP Collaboration** Collaboration, E. Komatsu *et al.*, "Five-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation," *Astrophys.J.Suppl.* **180** (2009) 330–376, arXiv:0803.0547 [astro-ph].

[188] **BOSS Collaboration** Collaboration, A. Ross *et al.*, "The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Analysis of potential systematics," *Mon.Not.Roy.Astron.Soc.* **424** (2012) 564, arXiv:1203.6499 [astro-ph.CO].

[189] L. Anderson, E. Aubourg, S. Bailey, F. Beutler, A. Bolton, *et al.*, "The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Measuring D_A and H at z=0.57 from the Baryon Acoustic Peak in the Data Release 9 Spectroscopic Galaxy Sample," arXiv:1303.4666 [astro-ph.CO].

[190] W. Valkenburg, "Perceiving the equation of state of Dark Energy while living in a Cold Spot," *JCAP* **1201** (2012) 047, arXiv:1106.6042 [astro-ph.CO].

[191] S. Alexander, T. Biswas, A. Notari, and D. Vaid, "Local Void vs Dark Energy: Confrontation with WMAP and Type Ia Supernovae," *JCAP* **0909** (2009) 025, arXiv:0712.0370 [astro-ph].

[192] T. Biswas, A. Notari, and W. Valkenburg, "Testing the Void against Cosmological data: fitting CMB, BAO, SN and H0," *JCAP* **1011** (2010) 030, arXiv:1007.3065 [astro-ph.CO].

[193] H. Alnes and M. Amarzguioui, "CMB anisotropies seen by an off-center observer in a spherically symmetric inhomogeneous Universe," *Phys.Rev.* **D74** (2006) 103520, arXiv:astro-ph/0607334 [astro-ph].

[194] C. Clarkson, T. Clifton, and S. February, "Perturbation Theory in Lemaitre-Tolman-Bondi Cosmology," *JCAP* **0906** (2009) 025, arXiv:0903.5040 [astro-ph.CO].

[195] S. February, C. Clarkson, and R. Maartens, "Galaxy correlations and the BAO in a void universe: structure formation as a test of the Copernican Principle," arXiv:1206.1602 [astro-ph.CO].

[196] J. Garcia-Bellido and T. Haugboelle, "The radial BAO scale and Cosmic Shear, a new observable for Inhomogeneous Cosmologies," *JCAP* **0909** (2009) 028, arXiv:0810.4939 [astro-ph].

[197] J. Garcia-Bellido and T. Haugboelle, "Looking the void in the eyes - the kSZ effect in LTB models," *JCAP* **0809** (2008) 016, arXiv:0807.1326 [astro-ph].

[198] K. Enqvist and T. Mattsson, "The effect of inhomogeneous expansion on the supernova observations," *JCAP* **0702** (2007) 019, arXiv:astro-ph/0609120 [astro-ph].

[199] A. Moss, J. Zibin, and D. Scott, "Precision Cosmology Defeats Void Models for Acceleration," *Phys.Rev.* **D83** (2011) 103515, arXiv:1007.3725 [astro-ph.CO].

[200] T. Clifton, P. Ferreira, and K. Land, "Living in a Void: Testing the Copernican Principle with Distant Supernovae," *Phys.Rev.Lett.* **101** (2008) 131302, arXiv:0807.1443 [astro-ph].

[201] M. Zumalacárregui, J. Garcia-Bellido, and P. Ruiz-Lapuente, "Tension in the Void: Cosmic Rulers Strain Inhomogeneous Cosmologies," *JCAP* **1210** (2012) 009, arXiv:1201.2790 [astro-ph.CO].

[202] W. Percival, S. Cole, D. Eisenstein, R. Nichol, J. Peacock, *et al.*, "Measuring the Baryon Acoustic Oscillation scale using the SDSS and 2dFGRS," *Mon.Not.Roy.Astron.Soc.* **381** (2007) 1053–1066, arXiv:0705.3323 [astro-ph].

[203] C. Clarkson and M. Regis, "The Cosmic Microwave Background in an Inhomogeneous Universe - why void models of dark energy are only weakly constrained by the CMB," *JCAP* **1102** (2011) 013, arXiv:1007.3443 [astro-ph.CO].

[204] S. Nadathur and S. Sarkar, "Reconciling the local void with the CMB," *Phys.Rev.* **D83** (2011) 063506, arXiv:1012.3460 [astro-ph.CO].

[205] A. Sandage, G. Tammann, A. Saha, B. Reindl, F. Macchetto, *et al.*, "The Hubble Constant: A Summary of the HST Program for the Luminosity Calibration of Type Ia Supernovae by Means of Cepheids," *Astrophys.J.* **653** (2006) 843–860, arXiv:astro-ph/0603647 [astro-ph].

[206] A. Riess, L. Macri, S. Casertano, H. Lampeitl, H. Ferguson, *et al.*, "A 3Telescope and Wide Field Camera 3," *Astrophys.J.* **730** (2011) 119, arXiv:1103.2976 [astro-ph.CO].

[207] W. Freedman, B. Madore, V. Scowcroft, C. Burns, A. Monson, *et al.*, "Carnegie Hubble Program: A Mid-Infrared Calibration of the Hubble Constant," *Astrophys.J.* **758** (2012) 24, arXiv:1208.3281 [astro-ph.CO].

[208] R. Sunyaev and Y. Zeldovich, "The Velocity of clusters of galaxies relative to the microwave background. The Possibility of its measurement," *Mon.Not.Roy.Astron.Soc.* **190** (1980) 413–420.

[209] C. Yoo, K. Nakao, and M. Sasaki, "CMB observations in LTB universes: Part II – the kSZ effect in an LTB universe," *JCAP* **1010** (2010) 011, arXiv:1008.0469 [astro-ph.CO].

[210] J. Zibin and A. Moss, "Linear kinetic Sunyaev-Zel'dovich effect and void models for acceleration," *Class.Quant.Grav.* **28** (2011) 164005, arXiv:1105.0909 [astro-ph.CO].

[211] P. Zhang and A. Stebbins, "Confirmation of the Copernican principle at Gpc radial scale and above from the kinetic Sunyaev Zel'dovich effect power spectrum," *Phys.Rev.Lett.* **107** (2011) 041301, arXiv:1009.3967 [astro-ph.CO].

[212] A. Raychaudhuri, "Relativistic cosmology. 1.," *Phys.Rev.* **98** (1955) 1123–1126.

[213] E. Poisson, *A Relativist's Toolkit*. Cambridge University Press, 2004.

[214] C. Quercellini, M. Quartin, and L. Amendola, "Possibility of Detecting Anisotropic Expansion of the Universe by Very Accurate Astrometry Measurements," *Phys.Rev.Lett.* **102** (2009) 151302, arXiv:0809.3675 [astro-ph].

[215] C. Quercellini, P. Cabella, L. Amendola, M. Quartin, and A. Balbi, "Cosmic Parallax as a probe of late time anisotropic expansion," *Phys.Rev.* **D80** (2009) 063527, arXiv:0905.4853 [astro-ph.CO].

[216] M. Quartin and L. Amendola, "Distinguishing Between Void Models and Dark Energy with Cosmic Parallax and Redshift Drift," *Phys.Rev.* **D81** (2010) 043522, arXiv:0909.4954 [astro-ph.CO].

[217] L. Eyer, B. Holl, D. Pourbaix, N. Mowlavi, C. Siopis, *et al.*, "The Gaia mission," arXiv:1303.0303 [astro-ph.IM].

[218] C. Quercellini, L. Amendola, A. Balbi, P. Cabella, and M. Quartin, "Real-time Cosmology," *Phys.Rept.* **521** (2012) 95–134, arXiv:1011.2646 [astro-ph.CO].

[219] J. Uzan, C. Clarkson, and G. Ellis, "Time drift of cosmological redshifts as a test of the Copernican principle," *Phys.Rev.Lett.* **100** (2008) 191303, arXiv:0801.0068 [astro-ph].

[220] C. Yoo, T. Kai, and K. Nakao, "Redshift Drift in LTB Void Universes," *Phys.Rev.* **D83** (2011) 043527, arXiv:1010.0091 [astro-ph.CO].

[221] B. de Swardt, P. Dunsby, and C. Clarkson, "Gravitational Lensing in Spherically Symmetric Spacetimes," arXiv:1002.2041 [gr-qc].

[222] H. Goto and H. Kodama, "Gravitational Lensing Effects in the LTB Model," arXiv:1002.3161 [astro-ph.CO].

[223] A. Linde, D. Linde, and A. Mezhlumian, "From the Big Bang theory to the theory of a stationary universe," *Phys.Rev.* **D49** (1994) 1783–1826, arXiv:gr-qc/9306035 [gr-qc].

[224] M. Cruz, M. Tucci, E. Martinez-Gonzalez, and P. Vielva, "The non-gaussian cold spot in wmap: significance, morphology and foreground contribution," *Mon.Not.Roy.Astron.Soc.* **369** (2006) 57–67, arXiv:astro-ph/0601427 [astro-ph].

[225] S. Foreman, A. Moss, J. Zibin, and D. Scott, "Spatial and temporal tuning in void models for acceleration," *Phys.Rev.* **D82** (2010) 103532, arXiv:1009.0273 [astro-ph.CO].

[226] A. Linde, D. Linde, and A. Mezhlumian, "Do we live in the center of the world?," *Phys.Lett.* **B345** (1995) 203–210, arXiv:hep-th/9411111 [hep-th].

[227] R. Sheth and G. Tormen, "Large scale bias and the peak background split," *Mon.Not.Roy.Astron.Soc.* **308** (1999) 119, arXiv:astro-ph/9901122 [astro-ph].

[228] J. Peacock, "Testing anthropic predictions for Lambda and the CMB temperature," *Mon.Not.Roy.Astron.Soc.* **379** (2007) 1067–1074, arXiv:0705.0898 [astro-ph].

[229] A. Knebe, S. Knollmann, S. Muldrew, F. Pearce, M. Aragon-Calvo, *et al.*, "Haloes gone MAD: The Halo-Finder Comparison Project," arXiv:1104.0949 [astro-ph.CO].

[230] M. Davis, G. Efstathiou, C. Frenk, and S. White, "The Evolution of Large Scale Structure in a Universe Dominated by Cold Dark Matter," *Astrophys.J.* **292** (1985) 371–394.

[231] S. Knollmann and A. Knebe, "Ahf: Amiga's Halo Finder," *Astrophys.J.Suppl.* **182** (2009) 608–624, arXiv:0904.3662 [astro-ph.CO].

[232] S. Patiri, A. Cuesta, F. Prada, J. Betancort-Rijo, and A. Klypin, "The alignment of dark matter halos with the cosmic web," *Astrophys.J.* **652** (2006) L75–L78, arXiv:astro-ph/0606415 [astro-ph].

[233] V. Marra, E. Kolb, S. Matarrese, and A. Riotto, "On cosmological observables in a swiss-cheese universe," *Phys.Rev.* **D76** (2007) 123004, arXiv:0708.3622 [astro-ph].

[234] T. Biswas and A. Notari, "Swiss-Cheese Inhomogeneous Cosmology and the Dark Energy Problem," *JCAP* **0806** (2008) 021, arXiv:astro-ph/0702555 [ASTRO-PH].

[235] V. Marra, E. Kolb, and S. Matarrese, "Light-cone averages in a swiss-cheese Universe," *Phys.Rev.* **D77** (2008) 023003, arXiv:0710.5505 [astro-ph].

[236] S. Weinberg, *Cosmology*. Oxford University Press, 2008.

[237] J. Peacock, "An introduction to the physics of cosmology,".

[238] K. Fisher, C. Scharf, and O. Lahav, "A Spherical Harmonic Approach to Redshift Distortion and a Measurement of Omega from the 1.2 Jy IRAS Redshift Survey," *Mon.Not.Roy.Astron.Soc.* **266** (1994) 219–226, arXiv:astro-ph/9309027 [astro-ph].

[239] **SDSS Collaboration** Collaboration, N. Padmanabhan *et al.*, "The Clustering of Luminous Red Galaxies in the Sloan Digital Sky Survey Imaging Data," *Mon.Not.Roy.Astron.Soc.* **378** (2007) 852–872, arXiv:astro-ph/0605302 [astro-ph].

[240] P. Peebles, *The Large-Scale Structure of the Universe*. Princeton University Press, 1980.

[241] M. Kerscher, I. Szapudi, and A. Szalay, "A Comparison of estimators for the two-point correlation function," *Astrophys.J.* **535** (2000) L13–L16, arXiv:astro-ph/9912088 [astro-ph].

[242] M. Vargas-Magana, J. Bautista, J.-C. Hamilton, N. Busca, E. Aubourg, *et al.*, "An optimized correlation function estimator for galaxy surveys," arXiv:1211.6211 [astro-ph.CO].

[243] H. Feldman, N. Kaiser, and J. Peacock, "Power spectrum analysis of three-dimensional redshift surveys," *Astrophys.J.* **426** (1994) 23–37, arXiv:astro-ph/9304022 [astro-ph].

[244] A. Moore, A. Connolly, C. Genovese, A. Gray, L. Grone, *et al.*, "Fast algorithms and efficient statistics: N-point correlation functions," arXiv:astro-ph/0012333 [astro-ph].

[245] R. Ponce, M. Cardenas-Montes, J. Rodriguez-Vazquez, E. Sanchez, and I. Sevilla, "Application of GPUs for the Calculation of Two Point Correlation Functions in Cosmology," arXiv:1204.6630 [astro-ph.IM].

[246] D. Bard, M. Bellis, M. Allen, H. Yepremyan, and J. Kratochvil, "Cosmological Calculations on the GPU," arXiv:1208.3658 [astro-ph.IM].

[247] D. Alonso, "CUTE solutions for two-point correlation functions from large cosmological datasets," arXiv:1210.1833 [astro-ph.IM].

[248] X. Xu, A. Cuesta, N. Padmanabhan, D. Eisenstein, and C. McBride, "Measuring D_A and H at z=0.35 from the SDSS DR7 LRGs using baryon acoustic oscillations," arXiv:1206.6732 [astro-ph.CO].

[249] X. Xu, N. Padmanabhan, D. Eisenstein, K. Mehta, and A. Cuesta, "A 2% Distance to z=0.35 by Reconstructing Baryon Acoustic Oscillations - II: Fitting Techniques," arXiv:1202.0091 [astro-ph.CO].